

-1501-

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4195> which encodes the amino acid sequence <SEQ ID 4196>. Analysis of this protein sequence reveals the following:

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1985(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 28 SFGTIRNSTALKQLTLDLNLSSFGTIRNSTALK 61

SFGTI+NS ALKQ + +N SFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKAQEEINQRSFGTIQNSIALK 40

Identities = 22/34 (64%), Positives = 26/34 (75%)

Query: 6 SFGTIRNSTALKLYAKQSPAFRSFGTIRNSTALK 39

SFGTI+NS ALK A++ RSFGTI+NS ALK

Sbjct: 7 SFGTIQNSIALKQKAQEEINQRSFGTIQNSIALK 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1372**

A DNA sequence (GBSx1457) was identified in *S.agalactiae* <SEQ ID 4197> which encodes the amino acid sequence <SEQ ID 4198>. Analysis of this protein sequence reveals the following:

Possible site: 16

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1407(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4199> which encodes the amino acid sequence <SEQ ID 4200>. Analysis of this protein sequence reveals the following:

Possible site: 37

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2055(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 154/221 (69%), Positives = 187/221 (83%)

Query: 1 MIKINFILDEPLVLSNATILTIEDVSVYSSLVKHFYQYDVDEHLKLFDDKQKSLKATEL 60

++ +NF +LDEP+ L TIL +EDV V+S +V++ YQY+ D LK FD K K++K +E+

-1502-

Sbjct: 8 LMNLNFSLLDEPIPLRGGTILVLEDVCVFSKIVQYCYQEEDSELKFFDHMKTIKSEI 67

Query: 61 MLVTDILGYDVNSAPILKLIHGDLENQFNEKPEVKSMVEKLAATITELIAFECLLENELDL 120  
MLVTDILG+DVNS+ ILKLIH DLE+QFNEKPEVKSM++KL ATITELI FECLLENELDL

5 Sbjct: 68 MLVTDILGFDVNSSTILKLIHADLESQFNEKPEVKSMIDKLVAITITELIVFECLLENELDL 127

Query: 121 EYDEIKILELIKALGVKIETQSDTIFEKCFEIIQVYHYLTKKNLLVFNNSGAYLTKDEVI 180  
EYDEI ILELIK+LGVK+ETQSDTIFEKC EI+Q++ YLTKK LL+FNNSGA+LTKDEV

10 Sbjct: 128 EYDEITILELIKSLGVKIVETQSDTIFEKCLEILQIFKYLTKKLLIFVNSGAFLTKDEVA 187

Query: 181 KLCEYINLMQKSVLFLEPRRLYDLPQYVIDKDYFLIGENMV 221  
L EYI+L +VLFLEPR LYD PQY++D+DYFLI +NMV

Sbjct: 188 SLQEYISLTNLTVLFLPRELYDFPQYILDEYFLITKNMV 228

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1373

A DNA sequence (GBSx1458) was identified in *S.agalactiae* <SEQ ID 4201> which encodes the amino acid sequence <SEQ ID 4202>. Analysis of this protein sequence reveals the following:

20 Possible site: 18  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.0842(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9783> which encodes amino acid sequence <SEQ ID 9784> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83918 GB:AL162753 hypothetical protein NMA0629 [Neisseria meningitidis Z2491]  
Identities = 45/104 (43%), Positives = 65/104 (62%), Gaps = 2/104 (1%)

35 Query: 4 RYMRMILMFDMPTETAERKAYRKFRKFLFLLSEGFIMHQFSVYSKLLNNTANNAMIGRLK 63  
++MR+I+ FD+P TA +RKA +FR+FLL +G+ M Q SVYS+++ + RL  
Sbjct: 5 KFMRIIVFFDLVPVITAARKKAANQFRQFLLKDGQYMLQLSVYSRIVKGRDSLQKHNNRLC 64

40 Query: 64 VNNPKKGNITLLTVTEKQFARMVYLHGERNT--SVANSDSRLVF 105  
N P++G+I L +TEKQ+A M L GE T NSD L+F  
Sbjct: 65 ANLPQEGSIRCLEITEKQYAAMKLLLGELKTQEKVNSDQLLLF 108

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4203> which encodes the amino acid sequence <SEQ ID 4204>. Analysis of this protein sequence reveals the following:

45 Possible site: 18  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.0822(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 97/112 (86%), Positives = 107/112 (94%)

55 Query: 1 MSYRYMRMILMFDMPTETAERKAYRKFRKFLFLLSEGFIMHQFSVYSKLLNNTANNAMIG 60  
MSYRYMRMILMFDMPT+TAEERKAYRKFRKFLFLLSEGFIMHQFS+YSKLLNNTANNAMIG

-1503-

Sbjct: 1 MSYRYMRMILMFDMPDTAEERKAYRKFRKFLLESGFIMHQFSIYSKLLNNTANNAMIG 60

Query: 61 RLKVVNNPKKGNITLLTVTEKQFARMVYLHGERNTSVANSDSRLVFLGDSYDQ 112  
 RL+ +NP KGNITLLTVTEKQFARM+YHGERN +ANSR RLVLFG+++D+

5 Sbjct: 61 RLREHNPKNKGNITLLTVTEKQFARMIVLHGERNNCIANSRDLVLFLGEAFDE 112

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1374

- 10 A DNA sequence (GBSx1459) was identified in *S.agalactiae* <SEQ ID 4205> which encodes the amino acid sequence <SEQ ID 4206>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB83919 GB:AL162753 hypothetical protein NMA0630 [Neisseria  
 meningitidis Z2491]  
 Identities = 71/224 (31%), Positives = 122/224 (53%)

- 25 Query: 4 WRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVDENI 63  
 WR++++ KLS + L+ + + + + L +I ++I+E + +++ L+ L +  
 Sbjct: 3 WRSLLIQNGGKLSLQRRQLLIQNGESHTVPLEDIAVIIENRETLLITAPLLSALAEHGA 62
- 30 Query: 64 LVIFCDDKRLPTAMLMPPYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFYLGE 123  
 ++ CD++ LP +PY H L Q++ E +K +W I+ QKILNQ+F E  
 Sbjct: 63 TLLTCDEQFLPCGQWLPYAQYHRQLKILKLQNLISEPLKKQLWQHIVRQKILNQAFVADE 122
- 35 Query: 124 CSFFEKSQSIMNLYHDLEFPDPSNREGHAARIYFNTLFGNDFSREQDNPINAGLDYGYSL 183  
 ++ + L ++ D NRE AA +YF LFG F+R +N +NA L+Y Y++  
 Sbjct: 123 TGNDLIAAKRLRTLASEVRSGDTGNREQAAALYFQALFGEKPTRNDNNNAVNAALNYTYAV 182
- 40 Query: 184 LLSMFAREVVKCGCMTQFGLKHANQFNQFNLASDIMEFFRPIVD 227  
 L + AR + G + GL H ++ N FNLA D +EP RP+ D  
 Sbjct: 183 LRAAVARALTLYGWLPLGLFHRSELNPFNLADDFIEPLRPLAD 226

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4207> which encodes the amino acid sequence <SEQ ID 4208>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

- 45 ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.3185(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 239/289 (82%), Positives = 271/289 (93%)

- 55 Query: 1 MAGWRTVVVNTHSKLSYKNNHLIFKDSYQTEMIHLSEIDILIMETTDIVLSTMLIKRLVD 60  
 MAGWRTVVVNTHSKLSYKNNHLIFKD+Y+TE+IHLSEIDIL++ETTDIVLSTML+KRLVD  
 Sbjct: 1 MAGWRTVVVNTHSKLSYKNNHLIFKDAYKTELIHLSEIDILLETTDIVLSTMLVKRLVD 60
- Query: 61 ENILVIFCDDKRLPTAMLMPPYARHDSSLQLSRQMSWIEDVKADVWTSIIAQKILNQSFY 120  
 EN+LVIFCDDKRLPTAMLMPP+Y RHDSSLQL +QMSW E VK+ VWT+IIAQKILNQ S Y

-1504-

Sbjct: 61 ENLVVIFCDDKRLPTAMLMPPFYGRHDSLQLGKQMSWSETVKSQVWTTIIAQKILNQSCY 120

Query: 121 LGECSFFEKSQSIMNLYHDLEPFDPSPNREGHAARIYFNTLFGNDFSREQDNPINAGLDYG 180  
 LG CS+FEKSQSIM+LYH LE FDPSPNREGHAARIYFNTLFGNDFS+ ++PINAGLDYG

Sbjct: 121 LGACSYFEKSQSIMDLYHGLENFDPSPNREGHAARIYFNTLFGNDFSRLDLEHPINAGLDYG 180

Query: 181 YSLLLSMFAREVVKCGCMTQFGLKHANQFNQFNLASDIMEFFRPPIVDRIIYENRQSDFKV 240  
 Y+LLLSMFAREVV GCMTQFGLKHANQFNQFN ASDIMEFFRP+VD+I+YENR F K

Sbjct: 181 YTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNASDIMEFFRPPLVDKIVYENRNQPFKP 240

Query: 241 MKRELFSMFSETYSYNGKEMYLSNIVSDYTKKVIKSLNSDNGIPEFRI 289  
 +KRELF++FS+T+SYNGKEMYL+NI+SDYTKK+K+LN++G G+PEFRI

Sbjct: 241 IKRELFTLFSDTFSYNGKEMYLTNIISDYTKKVVKALNNEGKGVPEFRI 289

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1375

A DNA sequence (GBSx1460) was identified in *S.agalactiae* <SEQ ID 4209> which encodes the amino acid sequence <SEQ ID 4210>. Analysis of this protein sequence reveals the following:

Possible site: 59  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1109(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB73943 GB:AL139078 hyopthetical protein Cj1523c [Campylobacter jejuni]  
 Identities = 165/746 (22%), Positives = 291/746 (38%), Gaps = 115/746 (15%)

Query: 318 LSASMIQRYDEHREDLKQLKQFVKASLPEKYQEI--FADSSKDGAGYIEGKTNQEAFFK 375  
 L+ S +R + L LK + Y++ F +S Y G + E ++

Sbjct: 50 LARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESIAKAYKGLISP--YELRFR 107

Query: 376 YLSKLLTKQEDSENFL--KIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRS----- 428  
 L++LL+KQ+ + L K + D ++ + G+I + E K + QS

Sbjct: 108 ALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIQNEEK-LANYQSVGEYL 166

Query: 429 --EYYPFLKENQDRIEKILTFRIPIY-----IGPLAREKSDFAW-MTRKTDDSI 474  
 EY+ KEN + + Y + + +++ +F + ++K ++ +

Sbjct: 167 YKEYFQKFKENSKEFTNVRNKKESYERCIASFLKDELKLIFFKQREFGFSFSKKFEEV 226

Query: 475 RPWNFEDLVDKEKSAEAFIHRMTNNDFFYLPEEKVLPKHSIYEKFTVYNELTKV--RYKN 532  
 F +++ + F H + N F+ +EK PK+S + F + + KN

Sbjct: 227 LSVAFY-----KRALKDFSHLVGNCSSFFT-DEKRAPKNSPLAFMFVALTRIINLNNLKN 280

Query: 533 EQGETYFFDSNIKQEIFDGVFKEHRKVS--KKLLDFLAKEYEFRIVDVIGLDKENKAF 590  
 +G Y D + + V K K KLL L+ +YE E +

Sbjct: 281 TEGILYTKDD--LNALLNEVLKNGTLTYKQTKKLLG-LSDDYE-----FKGEKGTY 328

Query: 591 NASLGTYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRENYKDLFTESQL 650  
 Y + K L + L D L +I + +TL +D +KK L Y ++Q+

Sbjct: 329 FIEFKKYKEFIKALGEHNSQDD----LNEIAKDITLIKDEIKLKKALAKYD--LNQNQI 382

Query: 651 KKLYRRHYTGWGRLSAKLINGIRDK--ESQKTILDYLDGSRNRFMQLINDDGLSFKS 708  
 L + + +S K + + E +K D+ + N IN+D F

Sbjct: 383 DSLSKLEFKDHLNISFKALKLVTPMLLEGKK-----YDEACNELNLKVAINEDKKDFLP 436

Query: 709 IISKAQAGSHSDNLKEVVGELAGSPAIIKKGILQSLKIVDELVKVMGYEPEQIVVEMAREN 768  
 ++ N P + + I + K+++ L+K G + +I +E+ARE

-1505-

Sbjct: 437 AFNETYYKDEVIN-----PVVLRAIKEYRKVLNALLKKYG-KVHKINIELAREV 484

Query: 769 QTTNQGR----RNSRQRYKLLDDG---VKNLASDLNG-NILKEYPTDNQALQNERLFLYY 820  
 + R + + YK D + L +N NILK L L+

5 Sbjct: 485 GKNHSQRAKIEKEQENENYKAKKDAELCEKLGKINSKNILK-----LRLFK 531

Query: 821 LQNGRDMYTGEALDIDNLSQ---YDIDHIIPQAFIKDSDINRVLVSSAKNRGKSDDVPS 877  
 Q Y+GE + I +L +IDHI P + DDS N+VLV + +N+ K + P

10 Sbjct: 532 EQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMKNKVLVFTKQKQEKLNQTP- 590

Query: 878 LEIVKDCKVFWKKL--LDAKLSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKH 935  
 E + W+K+ L L +++ L K ++ F R L +TR I +

Sbjct: 591 FFAFGNDSAKWQKIEVLAKNLPKKQKRILDK----NYKDKQKNFKDRNLNDTRYIARL 646

15 Query: 936 VARI-----LDERFNNELDSKGRIRKVKIVTLKSNLVSFRKEFGFYKIREVNYY 986  
 V L + N +L+ ++ KV + L S R +GF N+

Sbjct: 647 VLNYTKDYLDLPLSDDENTKLNDT-QKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHL 705

Query: 987 HHAHDAYLNAVVAKAILTKYPQLEPE 1012  
 HHA DA + A +I+ + + E

20 Sbjct: 706 HHAIIDAVIIAYANNSIVKAFSDFKKE 731

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4211> which encodes the amino acid sequence <SEQ ID 4212>. Analysis of this protein sequence reveals the following:

25 Possible site: 61  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0973(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 881/1380 (63%), Positives = 1088/1380 (78%), Gaps = 22/1380 (1%)

35 Query: 1 MNKPYSIGLDIGTNSVGSIIITDDYKVPAAKMRVLGNITDKEYIKKNLIGALLFDGGNTAA 60  
 M+K YSIGLDIGTNSVGW++ITD+YKVP+KK +VLGNITD+ IKKNLIGALLFD G TA  
 Sbjct: 1 MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKKFKVLGNITDRHSIKKNLIGALLFDSGETAE 60

40 Query: 61 DRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFHRLSDSFLVEEDKRGSKYPIFA 120  
 RLKRTARRRYTRR+NRI YLQEIF+ EM+KVDDSFHRLS+SFLVEEDK+ ++PIF  
 Sbjct: 61 ATRLKRTARRRYTRRNRIYLYLQEIFSNEMAKVDDSFHRLSDSFLVEEDKKHERHPHIFG 120

45 Query: 121 TLQEEKDYHEKFSTIYHLRKLADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNT 180  
 + +E YHEK+ TIYHLRK+L D +KADLRLIY+ALAH+IKFRGHFLIE D + N+

Sbjct: 121 NIVDEVAYHEKYPTIYHLRKKLVDSITDKADLRLIYALAHMIKFRGHFLIEGD-LNPDNS 179

Query: 181 DISKQYQDFLEIFNTTFFENNDLLSQNVDEAILTDKISKSAAKDRILAQYPNQKSTGIFA 240  
 D+ K + ++ +N FE N + + VD +AIL+ ++SKS + + ++AQ P +K G+F

50 Sbjct: 180 DVDKLFILQVQTYNQLFEEFNIPASGVDAKAILSARLSKSRLENLIAQLPGEKKNGLF 239

Query: 241 EFLKLIVGNQADFKKYFNLEDKTPQLQFAKDSYDEDLENLLGQIGDEFADLFSAACKLYDS 300  
 + L +G +FK F+L + LQ +KD+YD+DL+NLL QIGD++ADLF AAK L D+

55 Sbjct: 240 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLLAAKNLSDA 299

Query: 301 VLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDG 360  
 +LLS IL V TKAPLSASMI+RYDEH +DL LK V+ LPEKY+EIF D SK+GY

Sbjct: 300 ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFDQSKNGY 359

60 Query: 361 AGYIEGKTNQEAIFYKYLKLLTKQEDSENFLEIKKNEFLRKQRTFDNGSIPHQVHLTEL 420  
 AGYI+G +QE FYK++ +L K + +E L K+ ED LRKQRTFDNGSIPHQ+HL EL

Sbjct: 360 AGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL 419

Query: 421 KAIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDDSIIRPWNFE 480  
 AI+RRQ ++YPLK+N+++IEKILTFRIPYY+GPLAR S FAWMTRK++++I PWNFE

65

-1506-

Sbjct: 420 HAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAMWTRKSEETITPWNFE 479  
 Query: 481 DLVDKEKSAEAFIHRMTNNDFYLPPEEKVLPKHSLLIYEKFTVYNELTKVRYKNE-QGETYF 539  
 ++VDK SA++FI RMTN D LP EKVLPKHSLL+YE FTVYNELTKV+Y E + F  
 5 Sbjct: 480 EVVDKGASAQSFIERMTNFDKNLPNEKVLPHKHSLLYEYFTVYNELTKVKYVTEGMRKPAF 539  
 Query: 540 FDSNIKQEIFDGVFKEHRKVSKKLLDFLAKEEYEEFRIVDVIGLDKENKAFNASLGTYHD 599  
 K+ I D +FK +RKV+ K+L + K+ E F V++ G++ FNASLGTYHD  
 10 Sbjct: 540 LSGEQKKAIVDILLFKTNRKVTIVKQLKEDYFKKIECFDSVEISGVEDR---FNASLGTYHD 596  
 Query: 600 LEKIL-DKDFLDNPDNIESILEDIVQTLTLTFEDREMIKKRLNENYKDLFTESQLKKLYRRHY 658  
 L KI+ DKDFLDN +NE ILEDIV TLTLTFEDREMI++RL+ Y LF + +K+L RR Y  
 Sbjct: 597 LLKIIKDKDFLDNEENEDILEDIVLTLTLTFEDREMIERLKYAHLFDDKVMKQLKRRRY 656  
 15 Query: 659 TGWGRLSAKLINGIRDKESQKTIIDYLDGSRNRFMQNLINDGLSFKSIISKAQAGSH 718  
 TGWGRLS KLINGIRDK+S KTILD+L DG +NRNFMQLI+DD L+FK I KAQ  
 Sbjct: 657 TGWGRLSRKLINGIRDKQSGKTIIDFLKSDGFANRFMQNLIHDDSLTFKEDIQKAQVSGQ 716  
 Query: 719 SDNLKEVVGELAGSPAIIKKGILQSLKIVDELVKVMG-YEPEQIVVEMARENQTTNQGRRN 777  
 D+L E + LAGSPAIIKKGILQ++K+VDELVKVMG ++PE IV+EMARENQTT +G++N  
 20 Sbjct: 717 GDSLHEHIANLAGSPAIIKKGILQTVKVDDELVKVMGRHKPENIVEMARENQTTQKQKKN 776  
 Query: 778 SRQRYKLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEALDIDN 837  
 SR+R K +++G+K L S ILKE+P +N LQNE+L+LYYLQNGRDMY + LDI+  
 25 Sbjct: 777 SRERMKRIEKGELGS---QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINR 832  
 Query: 838 LSQYDIDHIIPQAFIKDSDIDNRVLSSAKNRGKSDDVPSLEIVKDCVFWKKLLDAKLM 897  
 LS YD+DHI+PQ+F+KDDSIDN+VL S KNRGKSD+VPS E+VK K +W++LL+AKL+  
 30 Sbjct: 833 LSDYDVIDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKNYWRQLLNAKLI 892  
 Query: 898 SQRYKYNLTAKAERGGLTSDDKARFIQRLVETROITKHVARILDERFNNELDSKGRRIK 957  
 +QRK+DNLTKAERGGL+ DKA FI+RQLVETROITKHVA+ILD R N + D + IR+  
 Sbjct: 893 TQRKFDNLTKAERGGLSELDKAGFIKQRLVETROITKHVAQILDSRMNTKYDENDKLIRE 952  
 35 Query: 958 VKIVTLKSNLVSFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGD 1017  
 VK++TLKS LVS+FRK+F FYK+RE+NNYHHAHDAYLNAV A++ KYP+LE EFVYGD  
 Sbjct: 953 VKVITLKSCLVDFRKDFQFYKIREVNNYHHAHDAYLNAVVGITALIKKYPKLESEFVYGD 1012  
 Query: 1018 YPKYN-----SYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEI 1070  
 Y Y+ S + AT K FFYSNIMNFFKT++TLA+G + + IE N +TGEI  
 40 Sbjct: 1013 YKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKPLIETNGETGEI 1072  
 Query: 1071 VWDKKKHAFATVRKVLSPQNNIVKKTETIQTGGFSKESILAHGNSDKLIPRKTCDIYLDPK 1130  
 VWDK + FATVRKVLSPQ NIVKKTET+QTGGFSKESIL NSDKLI RK KD DPK  
 45 Sbjct: 1073 VWDKGRDFATVRKVLSPQNNIVKKTETVQTGGFSKESILPKRNSDKLIARK-KD--WDPK 1129  
 Query: 1131 KYGGFDSPIVAVSVLVVADIKKGAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN 1190  
 KYGGFDSPI VAVSVLVVA ++KGK++KLK+V ELLGITIMERS FEKNP FLE+KGY  
 50 Sbjct: 1130 KYGGFDSPTVAVSVLVVAVKVEGKSKLKSVELLGITIMERSRFEKNPIDFLEAKGYKE 1189  
 Query: 1191 IRADKLIILPKYSLFELENGRRRLASAGELQKGNELALPTQFMKFLYLASRYNESKGP 1250  
 ++ D +I LPKYSLELENGR+R+LASAGELQKGNELALP++++ FLYLAS Y + KG P  
 Sbjct: 1190 VKKDLIIKLPKYSLELENGRRRLASAGELQKGNELALPSKYVNFLYLASHYEKLGSP 1249  
 55 Query: 1251 EEIEKKQEFVNQHSYFDDILQNLINDFSKRVILADANLEKINKLYQDNKENISVDELANN 1310  
 E+ E+KQ FV QH Y D+I++ I++FSKRVILADANL+K+ Y +++ + E A N  
 Sbjct: 1250 EDNEQKQLFVEQHKHYLDEIEQISEFSKRVILADANLDKVL SAYNKHHRDK-PIREQAEN 1308  
 Query: 1311 IINLFTFTSLGAPAAFKFDDKIIVDRKRYTSTKEVLNSTLIHQISITGLYETRIDLKLGED 1370  
 II+LFT T+LGAPAAFK+FD +DRKRYTSTKEVL++TLIHQISITGLYETRIDL +LG D  
 60 Sbjct: 1309 IIHFLFTLNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQISITGLYETRIDLSQLGGD 1368

SEQ ID 4210 (GBS317) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell  
 extract is shown in Figure 27 (lane 2; MW 179.3kDa) and in Figure 159 (lane 5 & 6; MW 180kDa). It was  
 65 also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure  
 27 (lane 3; MW 154.3kDa) and in Figure 159 (lane 9 & 10; MW 154kDa).

-1507-

GBS317-GST was purified as shown in Figure 224, lane 9-10. GBS317-His was purified as shown in Figure 222, lane 9.

GBS317N was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 149 (lane 2-4; MW 116kDa).

5 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 166 (lane 6-8; MW 92kDa).

GBS317dN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 187 (lane 7; MW 116kDa). Purified GBS317dN-GST is shown in Figure 245, lane 8.

10 GBS317C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 13; MW 92kDa). Purified GBS317dC-GST is shown in Figure 245, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1376

15 A DNA sequence (GBSx1461) was identified in *S.agalactiae* <SEQ ID 4213> which encodes the amino acid sequence <SEQ ID 4214>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -11.94	Transmembrane	132 - 148 ( 123 - 156)
INTEGRAL	Likelihood = -11.09	Transmembrane	190 - 206 ( 183 - 209)
INTEGRAL	Likelihood = -4.94	Transmembrane	95 - 111 ( 94 - 115)

----- Final Results -----

bacterial membrane	---	Certainty=0.5776(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related sequence was also identified in GAS <SEQ ID 9133> which encodes the amino acid sequence <SEQ ID 9134>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -7.32	Transmembrane	126 - 142
INTEGRAL	Likelihood = -6.90	Transmembrane	178 - 194

----- Final Results -----

bacterial membrane	---	Certainty=0.3930(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 5	LMKDKLLVVLTVIWIISLATLATIYIAWLIYPIEIQFLKLEKVVYLKAETIYYNFNKLMI 64
45	+M + ++ +W+W+++LA L TIY WL YP+E+ LKLE+VV++ + I +N+N L+
Sbjct: 4	VMVENTKLLCSWVWLLALAILITIYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLLN 63
Query: 65	YLTHPFISDLNMPSEFPSSDGLKHFADVKYLFTHLHGLFVILTFPVIYFLRRGWKQKSIF 124
50	YLT+PF++ L +F SS DGLKHFADVK+LF L +F+ L +P + + K K +
Sbjct: 64	YLTNPFVTRLEFANFHSADGLKHFADVKWLFHLTQVVFLLGLLYPTLKTFTQRLKTKRFW 123

-1508-

Query: 125 LYEGFFKIAIMLPFIVVCAFLLGFDQFFTLFHEVLFPDSTWQFNPLTDPVIWILPETF 184  
 L + +A + P+ I + A +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F  
 Sbjet: 124 LLQKPLILAAALFPLMIGLMASFIGFEHFFTLFHQVLFVGDSSWLFDPDKDSVIWILPEVF 183

Query: 185 FLHCFIIFLLIYETITITILLIIGR 208  
 FLHCF+ F+++YE I L+ + R  
 Sbjet: 184 FLHCFLFFMIVYEIILWSLVGLAR 207

10 SEQ ID 4214 (GBS167) was expressed in and purified from *E.coli*. The purified protein is shown in lanes 5 & 6 of Figure 223.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1377

15 A DNA sequence (GBSx1462) was identified in *S.agalactiae* <SEQ ID 4217> which encodes the amino acid sequence <SEQ ID 4218>. This protein is predicted to be p-nitrophenyl phosphatase (pho2). Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3925(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism  
 [Bacillus subtilis]  
 Identities = 121/249 (48%), Positives = 172/249 (68%)

30 Query: 3 YKGYLIDLDGTIYKGSRIPIAGERFIERLQEKGIPLYMLVTNNTTRTPESVQEMLRGFNVE 62  
 YKGYLIDLDGT+Y G +I F+ L+++G+PY+ VTNN++RTP+ V + L F++  
 Sbjet: 4 YKGYLIDLDGTYNGTEKIEEACEFVRTLKDRGVPYLFVTNNSRTPKQVADKLVSFDIP 63

35 Query: 63 TPLETIYTATMATVDYMNDMNRGKTAYVIGEEGLKKAIDAGYVEDTKNPAYVVVGLDWN 122  
 E ++T +MAT ++ + + YVIGEEG+++AI + G +N +VVVG+D +  
 Sbjet: 64 ATEEQVFTTSMATAQHIAQQKKDASVYVIGEEGIRQAIEENGLTFGGENADFFVVVGIDRS 123

40 Query: 123 VTYDKLATATLAIQNGALFIGTNPDLNIPTERGLLPAGSLNALLEAATRIKPVFIGKPN 182  
 +TY+K A LAI+NGA FI TN D+ IPTERGLLP GSL ++L +T ++PVFIGKP  
 Sbjet: 124 ITYEKFAVGCLAIRNGARFISTNGDIAIPTERGLLPNGSLTSLTVSTGVQPVFIGKPE 183

45 Query: 183 AIIMNKALEILNIPRQAVMVGDNLYLTDIMAGINNDITLLVTTGFTTVEEVPDLPIQPS 242  
 +IIM +A+ +L ++ +MVGDN YTDIMAGIN +DTLLV TG T E + D +P+  
 Sbjet: 184 SIIMEQAMRVLGTDVSETLMVGDNATDIMAGINAGMDTLLVHTGVTKREHMTDDMEKPT 243

Query: 243 YVLASLDEW 251  
 + + SL EW  
 Sbjet: 244 HAIDSLTEW 252

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4219> which encodes the amino acid sequence <SEQ ID 4220>. Analysis of this protein sequence reveals the following:

Possible site: 48  
 >>> Seems to have no N-terminal signal sequence

55 INTEGRAL Likelihood = -0.53 Transmembrane 128 - 144 ( 128 - 144)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1213(Affirmative) < succ>

-1509-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAB15219 GB:Z99120 similar to N-acetyl-glucosamine catabolism  
 [Bacillus subtilis]  
 Identities = 121/250 (48%), Positives = 166/250 (66%), Gaps = 1/250 (0%)

10 Query: 3 YKGYLIDLDGTIYQGNRIPAGERFIKRLQERGIPYLLVTNNTRTPPEMVQSMANQFHV 62  
 YKGYLIDLDGT+Y G +I F++ L++RG+PYL VINN++RTP+ V L + F +  
 Sbjet: 4 YKGYLIDLDGTMNGTEKIEEACEFVRITLKDRGVYLYFVTNNSSRTPKQVADKLVS-FDI 62

15 Query: 63 ETSIETIYTATMATVDYMDNMNRGKTAYVIGETGLKSAIAAGYVEELENPAYVVVGLDS 122  
 + E ++T +MAT ++ + + YVIGE G++ AI G EN +VVVG+D  
 Sbjet: 63 PATEEQVFTTSMATAQHIAQKKDASVYVIGEEGIRQAIEENGLTFGGENADFVVVGIDR 122

20 Query: 123 QVTYEMLAIALTLAIQKALFIGTNPDLNIPTEGLMPGAGALNALLEAATRVKPVFIGKP 182  
 +TYE A+ LAI+ GA FI TN D+ IPTEGL+PG G+L ++L +T V+PVFIGKP  
 Sbjet: 123 SITYEKFAVGCLAIRNGARFISTNGDIAIPTEGLLPNGSLTSVLTVSTGVQPVFIGKP 182

25 Query: 183 NAIIMNKSLEVLGIQRSEAVMVGDNLYLTDIMAGI QNDIATILVTTGFTTRPEEVPDLPIQP 242  
 +IIM +++ VLG SE +MVGDN YTDIMAGI + T+LV TG T+ E + +P  
 Sbjet: 183 ESIIMEQAMRVLGTDVSETLMVGDN YATDIMAGINAGMDTLLVHTGVTKREHMTDDMEKP 242

Query: 243 DHVLSLDEW 252  
 H + SL EW  
 Sbjet: 243 THAIDSLTEW 252

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 207/250 (82%), Positives = 227/250 (90%), Gaps = 1/250 (0%)

Query: 3 YKGYLIDLDGTIYKGSRI PAGERFIERLQEKGI PYMLVTNNTRTPESVQEMLRG-FNV 61  
 YKGYLIDLDGTIY+GK+RIPAGERFI+RLQE+GIPY+LVTNNTRTPE VQ ML F+V  
 35 Sbjet: 3 YKGYLIDLDGTIYQGNRIPAGERFIKRLQERGIPYLLVTNNTRTPPEMVQSMANQFHV 62

Query: 62 ETPLETIYTATMATVDYMDNMNRGKTAYVIGEEGLKKAIDAGYVEDTKNPAYVVVGLDW 121  
 ET +ETIYTATMATVDYMDNMNRGKTAYVIGE GLK AIA AGYVE+ +NPAYVVVGLD  
 40 Sbjet: 63 ETSIETIYTATMATVDYMDNMNRGKTAYVIGETGLKSAIAAGYVEELENPAYVVVGLDS 122

Query: 122 NVTYDKLATATLAIQNGALFIGTNPDLNIPTEGLLPAGAGSLNALLEAATRIKPVFIGKP 181  
 VTY+ LA ATLAIQ GALFIGTNPDLNIPTEGL+PGAG+LNALLEAATR+KPVFIGKP  
 45 Sbjet: 123 QVTYEMLAIALTLAIQKALFIGTNPDLNIPTEGLMPGAGALNALLEAATRVKPVFIGKP 182

Query: 182 NAIIMNKALEILNIPRNQAVMVGDNLYLTDIMAGINNDIDTLLVTTGFTTVEEVPDLPIQP 241  
 NAIIMNK+LE+L I R++AVMVGDNLYLTDIMAGI NDI T+LVTTGFT EEVP LPIQP  
 50 Sbjet: 183 NAIIMNKSLEVLGIQRSEAVMVGDNLYLTDIMAGI QNDIATILVTTGFTTRPEEVPDLPIQP 242

Query: 242 SYVLASLDEW 251  
 +VL+SLDEW  
 Sbjet: 243 DHVLSLDEW 252

A similar DNA sequence was identified in *S.pyogenes* <SEQ ID 4215> which encodes amino acid sequence <SEQ ID 4216>. An alignment of the GAS and GBS sequences follows:

55 Identities = 94/204 (46%), Positives = 139/204 (68%)

Query: 4 VMVENTKLLCSWVWLLAILITIIYSTWLWYPLEVDHLKLEQVVFMSKDAILHNYNGLIN 63  
 +M + ++ +W+W+++LA L TIY WL YP+E+ LKLE+VV++ + I +N+N L+  
 60 Sbjet: 5 LMKDKLLVVLTIWIISLATLATIYIAWLIYPIETIQFLKLEKVVYLKAETIYYNFNKLMI 64

Query: 64 YLTNPFVTRLEFANFHSSADGLKHFADVKWLFHLTQVVFLGLLYPTLKTFTQRLKTKRFW 123  
 YLT+PF++ L +F SS DGLKHFADVK+LF L +F+ L +P + + K K +  
 Sbjet: 65 YLTHPFIISDLNMPSPFSSDGLKHFADVKYLFTHAGLFLVILTFPVIYFLRRGWKQKSIF 124

Query: 124 LLQKPLILAALEFPLMIGLMASFIGFEHFFTLFHQVLFVGDSSWLFDPDKDSVIWILPEVF 183

-1510-

L + +A + P+ I + A +GF+ FFTLFH+VLF GDS+W F+PL D VIWILPE F  
 Sbjct: 125 LYEGFFKIAIMLPFIVVCAFLLGFDQFFTLFHEVLFPDSTWQFNPLTDPVIWILPETF 184

Query: 184 FLHCFLLFFMIVYEIILWLSLVGLAR 207

FLHCF+ F+++YE I L+ + R

Sbjct: 185 FLHCFIIFLLIYETITILLIIGR 208

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 10 Example 1378

A DNA sequence (GBSx1463) was identified in *S.agalactiae* <SEQ ID 4221> which encodes the amino acid sequence <SEQ ID 4222>. This protein is predicted to be oleoyl-acyl carrier protein thioesterase. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3332(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB02069 GB:AB026647 acyl carrier protein thioesterase

[Arabidopsis thaliana]

Identities = 59/248 (23%), Positives = 104/248 (41%), Gaps = 30/248 (12%)

Query: 2 GLLYRETYEVPFYESDTNHYMKLPQLLALALQISAKQSLKLGIGDD-----IVFKRYGLV 56

GL Y+E + V YE +N + + L ++ + +G D ++ L+

Sbjct: 81 GLSYKEKFVRSYEVGSNKTATVETIANLLQEVGCNHAQSVGFSTDGFATTTTMRKLHLI 140

Query: 57 WVVTDYIIDIERLPKHAEKIVIEATEAHNKLCCYRYFYIYGE-DGQKIITISSAFVLM 115

WV I+I + P + + IET ++ ++ R + + G+ +S +V+M+

Sbjct: 141 WVTARMHIEIYKYPAGDGVVEIETWCQSEGRIGTRRDWILKDSVTGEVTGRATSKWVMN 200

Query: 116 FKTRKIHPVLDDITSY-----QSQRICKVIRGPKYHPIGDSKVKQYHVR 160

TR++ V DD+ Y ++ +KK+ PK + R

Sbjct: 201 QDTRRLQKVSDDVRDEYLVFCPQEPRLAFPEENNRSLKKI---PKLEDPAQYSMIGLKPR 257

Query: 161 YFDLDMNGHVNSKYLEWMDVLDLDFLSSHIPKKIDLYKEIQYGTDIKSHWYQDGLV 220

DLDMN HVNN Y+ W+ + + + +H + I L Y +E Q + D L

Sbjct: 258 RADLDMNQHVNNVTYIGWVLESIPQEIVDTHELQVITLDYRRECQDDVV-----DSL 311

Query: 221 TRHDIIGG 228

T IGG

Sbjct: 312 TTTSEIGG 319

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4223> which encodes the amino acid sequence <SEQ ID 4224>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.88 Transmembrane 21 - 37 ( 21 - 38)

----- Final Results -----

bacterial membrane --- Certainty=0.2550(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1511-

>GP:AAB71730 GB:U65643 acyl-ACP thioesterase [*Myristica fragrans*]  
Identities = 41/128 (32%), Positives = 67/128 (52%), Gaps = 11/128 (8%)

Query: 33 FIFMIKRGGLLDILAYFALLNPDTRKVTIPEDLVAPFETDFVKKLHRV-----PKMPL 87  
F+ K G +L + + ++N TR+++ IPE++ E FV+ H V K+P  
Sbjct: 147 FLRDCKTGEILTRATSVWVMNKRTRRLSKIPEEVRVEIEPYFVE--HGVLDDEDSRKLKP 204

Query: 88 LEQS----IDRDYVRYFDIDMNGHVNNISKYLDWMYDVLGCEFLKTHQPLKMTLKYYKEV 143  
L + I R R+ D+D+N HVNN KY+ W+ + + L++H+ MTL+Y KE  
Sbjct: 205 LNDNTANYIRRLAPRWSDLDVNQHVNNVKYIGWILESVPSLLSHELYGMTLEYRKEC 264

Query: 144 SPGGQITS 151  
G + S  
Sbjct: 265 GKDGLLQS 272

An alignment of the GAS and GBS proteins is shown below.

Identities = 62/144 (43%), Positives = 94/144 (65%)

Query: 101 GQKIITISSAFVLMDFKTRKIHPVLDDITSYQSQRICKVIRGPKYHPIGDSKVKQYHVR 160  
G ++ I + F L++ TRK+ + +D+ + +++ +KK+ R PK + S + Y+VR  
Sbjct: 40 GGLLVDILAYFALLNPDTRKVTIPEDLVAPFETDFVKKLHRVFKMPLLEQSIDRDYVVR 99

Query: 161 YFDLDMNGHVNNISKYLEWMYDVLDDLFLSSHIPKKIDLKYYKEIQYGTDIKSHWYQDGLV 220  
YFD+DMNGHVNNISKYL+WMYDVL +FL +H P K+ LKY+KE+ G I S ++ D L  
Sbjct: 100 YFDIDMNGHVNNISKYLDWMYDVLGCEFLKTHQPLKMTLKYYKEVSPGGQITSSYHLDQLT 159

Query: 221 TRHDIIGGDAIHAQARIEWQEKKE 244  
+ H I ++AQA IEW+ K+  
Sbjct: 160 SYHQITSDGQLNAQAMIEWRAIKQ 183

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1379

A DNA sequence (GBSx1464) was identified in *S. agalactiae* <SEQ ID 4225> which encodes the amino acid sequence <SEQ ID 4226>. This protein is predicted to be coproporphyrinogen III oxidase. Analysis of this protein sequence reveals the following:

Possible site: 40  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05062 GB:AP001511 coproporphyrinogen III oxidase [*Bacillus halodurans*]  
Identities = 173/375 (46%), Positives = 248/375 (66%), Gaps = 5/375 (1%)

Query: 5 PTSAYVHIPFCTQICYCDFSKVFIKNQPVDAYLQALIREFR----SYDITELRTLYIGG 60  
P +AY+HIPFC ICYCDF+K ++KNQPV+ YLQAL E L+TLY+GG  
Sbjct: 2 PKAAYIHIPFCEHICYCDFNKFYLNQPVNEYLQALETEMAMVVAEQPTKSLQTLTYVGG 61

Query: 61 GTPTSISAVQLDYLLTELSRDNLNLTLEEFTEANPGDLTVDKIEVLQKSAVNRVSLGVQ 120  
GTPT+++A QL LL + R L L+ LEEFT E NP + +K++VL+ V+R+S+GVQ  
Sbjct: 62 GTPTALTADQLAQLLASIKRTLPLSDLEEFFEVPNPDSEDEKLDVLRSGVDRLSIGVQ 121

Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNISIDLIYALPGQTMDDVRSNVAKALSL 180  
F LK IGR+H++ + ++ + AGF N+S+DL+ LP QT + + +A +L  
Sbjct: 122 AFQPLLLKEIGRTHDQKSVEQAVEKSRQAGFANLSLDLMLGLPKQTPMFAETLKEAFAL 181

-1512-

Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKLHLPTEDLEAEMFEYIISEMERNGFEHYEISNFT 240  
 + HLS YSL +E TVF N+ R+G+L LP ED E +M+ + E E++GF+ YEISNF  
 Sbjct: 182 EVEHLSCYSLKVEAKTVFYNRQGRITLPPEDDEVKMYRQLCYETEKHGFKQYEISNFA 241

5 Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSSEGNARLSE-EVL 299  
 K G+ESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSSEGNARLSE-EVL 299  
 Sbjct: 242 KKGYESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSSEGNARLSE-EVL 301

10 Query: 300 SKNEMMEEELFLGLRKKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMT 359  
 S+ E MEE++FLGLRK+ GV F ++FG S Y + + +L+ LL+ + +++T  
 Sbjct: 302 SRVEQMEEQMFLGLRKRSGVEERVVFVERFGVSMFSLYEKQIAQLVARCLLERTDDRRLT 361

Query: 360 KKGLFLGDTVAEKFI 374  
 +GL LG+ V E+F+

15 Sbjct: 362 DEGLLLGNEVFQFL 376

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4227> which encodes the amino acid sequence <SEQ ID 4228>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3202(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 304/376 (80%), Positives = 343/376 (90%)

30 Query: 1 MLKKPTSAYVHIPFCTQICYYCDFSKVFIKNQPVDAYLQALIREFRSYDITELRTLYIGG 60  
 M KKPTSAYVHIPFCTQICYYCDFSKVFI+NQPVDAYL+ALI+EF SY I +L+TLYIGG  
 Sbjct: 33 MSKKPTSAYVHIPFCTQICYYCDFSKVFIQNQPVDAYLKALIQEFD SYGIRD LKTLYIGG 92

35 Query: 61 GTPTSISAVQLDYLLTELSDRLNLTLEEFTEANPGDLTVDKIEVLQKSAVNRVSLGVQ 120  
 GTPT+I+A QL+YLL L R+LNL+ LEEFTIEANPGDLT +KI VLQ+SAVNR+SLGVQ  
 Sbjct: 93 GTPTAITAKQLEYLLNHLERNLNLDDLEEFTEANPGDLTPEKIAVLQSAVNRISLGVQ 152

40 Query: 121 TFNDKHLKRIGRSHNEAQIYSTIDALKTAGFQNI SIDLIYALPGQTMDDVRSNVAKALSL 180  
 TFN+K LK+IGRSHNE QIYSTI LKTAGF NISIDLIYALPGQT+D V+ NVAKAL+L  
 Sbjct: 153 TFNNKQLKQIGRSHNEEQIYSTIANLKTAGFHNISIDLIYALPGQTLQVQKENVAKALAL 212

45 Query: 181 NIPHLSLYSLILEHHTVFMNKMRRGKLHLPTEDLEAEMFEYIISEMERNGFEHYEISNFT 240  
 +IPHLSLYSLILEHHTVFMNKMRRGKL+LPTEDLEAEMFEYIISEME NGFEHYEISNFT  
 Sbjct: 213 DIPHLSLYSLILEHHTVFMNKMRRGKLNLPTEDLEAEMFEYIISEMEANGFEHYEISNFT 272

Query: 241 KPGFESRHNLMYWDNVEYYGVGAGASGYLDGIRYRNRGPIQHYLKGVSSEGNARLSEEVLS 300  
 KPGFESRHNLMYWDNVEY+G GAGASGYL+GIRY+NR PIQHYLK V GNARL+EEVL  
 Sbjct: 273 KPGFESRHNLMYWDNVEYFGCGAGASGYLNGIRYQNRVPIQHYLKAVEAGNARLNEEVLR 332

50 Query: 301 KNEEMEEELFLGLRKKKEGVSIGKFEQKFGTSFEKRYGQIVQELQSDGLLKENNGFIQMTK 360  
 K EMMEEELFLGLRKK GVSI +F++KFG SFE+RYG IV+ELQ+ GLL +++ F++MTK  
 Sbjct: 333 KEEMMEEELFLGLRKKTGVSIGKFEQKFGMSFEERYGNIVRELQNGLLVKDDAFVRMTK 392

55 Query: 361 KGLFLGDTVAEKFI 376  
 KGLFLGD+VAE+FI++  
 Sbjct: 393 KGLFLGDSVAERFILD 408

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1513-

**Example 1380**

A DNA sequence (GBSx1465) was identified in *S.agalactiae* <SEQ ID 4229> which encodes the amino acid sequence <SEQ ID 4230>. Analysis of this protein sequence reveals the following:

```

Possible site: 52
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3729(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1381**

A DNA sequence (GBSx1466) was identified in *S.agalactiae* <SEQ ID 4231> which encodes the amino acid sequence <SEQ ID 4232>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2989(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
25     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4233> which encodes the amino acid sequence <SEQ ID 4234>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2993(Affirmative) < succ>
35     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 36/109 (33%), Positives = 58/109 (53%), Gaps = 6/109 (5%)
40
Query: 9  WAKHKYLVLSKSQKIYLDIRQTLKSPNCT---VLDVQSLIDQAVLLEESPSQVTNAYMH1 65
      WA  KY V++ SQ+ Y +R+ K   + VL   LI++A + + + AY H+
Sbjct: 13 WAYQKYVWMAHSQQHYNALRELFGKNQWSEEKVLTFHCLIEEAQAIPPTVKSLRTAYQHV 72

Query: 66 WGYFKNKAERQEEFLTLLEKYRKTGYQRRKLLAFLKQLLAKYPNSYL 114
      WGYFK A ++EK+ F L +   + ++L FL+++ A Y SYL
Sbjct: 73 WGYFKKVASQEEKDHFKDLDQALET---KSEEMLCFLQEMTAHYQPSYL 118

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1514-

**Example 1382**

A DNA sequence (GBSx1467) was identified in *S.agalactiae* <SEQ ID 4235> which encodes the amino acid sequence <SEQ ID 4236>. This protein is predicted to be mrsA (mrsA). Analysis of this protein sequence reveals the following:

```

5      Possible site: 35
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.96      Transmembrane      56 - 72 ( 56 - 72)

      ----- Final Results -----
10      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15      >GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)
      [Bacillus subtilis]
      Identities = 284/451 (62%), Positives = 353/451 (77%), Gaps = 4/451 (0%)

      Query: 1      MGKYFGTGDGVRGEANVELTPELAFKLRFGGYVLSQHETDRPRVFVARDTRISGEMLESA 60
20      MGKYFGTGDGVRG AN ELTPELAFK+GRFGGYVL++ + RP+V + RDTRISG MLE A
      Sbjct: 1      MGKYFGTGDGVRGVANSELTPELAFKVRFGGYVLTK-DKQRPKVLIQRDTRISGHMLEGA 59

      Query: 61      LIAGLLSVGIEVYKLGVLATPGVSYLVRTEKASAGVMISASHNPALDNGIKFFGSDGFKL 120
25      L+AGLLS+G EV +LGV++TPGVSYL + A AGVMISASHNP DNGIKFFG DGFKL
      Sbjct: 60      LVAGLLSIGAEVMRLGVISTPGVSYLTKAMDAEAGVMISASHNPVQDNGIKFFGGDGFKL 119

      Query: 121     DDDRELEIEALLDAKEDTLPRPSAQGLGTLVDYPEGLRKYEFMESTGI-DLEGMKVALD 179
30      D++E EIE L+D ED LPRP LG + DY EG +KY +F++ T D G+ VALD
      Sbjct: 120     SDEQAEIERLMDEPEDKLPFVGADLGLVNDYFEGGQKYLQFLKQTADEFTGIHVALD 179

      Query: 180     TANGAATASARNIFLDLADISVIGDQPDGLNINDGVGSTHPEQLQSLVRENGSDIGLAF 239
35      ANGA ++ A ++F DL+AD+S +G P+GLNINDGVGSTHPE L + V+E +D+GLAF
      Sbjct: 180     CANGATSSLATHLFDADLADVSTMGTSFNGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

      Query: 240     DGDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQNITVTTVMSNLGFHKALDREGI 299
40      DGD DRLIAVDE G IVDGD+IM+I K+L +G+L +T+V+TVMSNLGF+KAL++EGI
      Sbjct: 240     DGDGDRLIAVDEKGNIVDGDQIMYICSKHLKSEGRKDDTVVSTVMSNLGFYKALEKEGI 299

      Query: 300     HKAITAVGDRYVVEEMRKSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKL 359
45      TAVGDRYVVE M+K GYN+GGEQSGH+I +DYNTTGDG L+AI L +K TGK L
      Sbjct: 300     KSVQTAVGDRYVVEAMKKGYNVVGGEQSGHLIFLDYNTTGDGLLSAIMLMNTLKATGKPL 359

      Query: 360     SELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEEMDGNRILVRPSGTEPL 419
50      SELA+E+ +PQ LVN+RV + K K E + +I+++E+EM+G+GRILVRPSGTEPL
      Sbjct: 360     SELAAEMQKFPQLLVNVRVTD--KYKVEENEKVKAVISEVEKEMNGDGRILVRPSGTEPL 417

      Query: 420     LRVMAEAPTNEAVDYYVDTIADVVRTEIGLD 450
55      +RVMAEA T E D YV+ I +VVR+E+GL+
      Sbjct: 418     VRVMAEAKTKELCDEYVNRIVEVVRSEMGL 448

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4237> which encodes the amino acid sequence <SEQ ID 4238>. Analysis of this protein sequence reveals the following:

```

55      Possible site: 35
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -0.96      Transmembrane      56 - 72 ( 56 - 72)

      ----- Final Results -----
60      bacterial membrane --- Certainty=0.1383(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1515-

The protein has homology with the following sequences in the databases:

>GP:CAB11970 GB:Z99105 similar to phosphoglucomutase (glycolysis)

[Bacillus subtilis]

Identities = 287/451 (63%), Positives = 346/451 (76%), Gaps = 4/451 (0%)

5  
Query: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETERPKVVFVARDTRISGEMLESA 60  
MGKYFGTDGVRG AN ELTPELAFK+GRFGGYVL++ + +RPKV + RDTRISG MLE A  
Sbjct: 1 MGKYFGTDGVRGVANSELTPPELAFKVGFRGGYVLTG-DKQRPKVLIGRDTRISGHMLEGA 59

10  
Query: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGNDGFKL 120  
L+AGLLS+G EV +LGV++TPGVSYL + A AGVMISASHNP DNGIKFFG DGFKL  
Sbjct: 60 LVAGLLSIGAEVMRLGVISTPGVSYLTAMDAEAGVMISASHNPVQDNGIKFFGGDGFKL 119

15  
Query: 121 ADDQELEIEALLDAPEDTLPRPSAEGGLTLVDYPEGLRKYEFVLTGT-DLSGMTVALD 179  
+D+QE EIE L+D PED LPRP LG + DY EG +KY +FL T D +G+ VALD  
Sbjct: 120 SDEQEAETIERLMDPEDEKLPKRVGADLGLVNDYFEGGQKYLQFLKQTADEFTGIHVALD 179

20  
Query: 180 TANGAASVSARDVFLDLNABIAVIGKEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAF 239  
ANGA S A +F DL+A+++ +G PGNLNINDGVGST PE L VKE ADLGLAF  
Sbjct: 180 CANGATSSLATHLEFADLDADVSTMGTSPEGLNINDGVGSTHPEALSAFVKEKNADLGLAF 239

25  
Query: 240 DGDSDRLIAVDDEGEIVDGDRIIMFIIGKYLSEKGLLAHNTIVTTVMSNLGFHKALDKQGI 299  
DGD DRLIAVDE G IVDGD+IM+I K+L +G L +T+V+TVMSNLGF+KAL+K+GI  
Sbjct: 240 DGDGDRLIAVDEKGNIVDGDQIMYICSHKLGSEGLKDDTVVSTVMSNLGFYKALEKEGI 299

30  
Query: 300 NKAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKSL 359  
TAVGDRYVVE M+ GYN+GGEQSGH+I +DYNTTGDG L+AI L +K TGK L  
Sbjct: 300 KSVQTAVGDRYVVEAMKKDGYNVGGEQSGHLIFLDYNTTGDGLLSAIMLMNTLKATGKPL 359

35  
Query: 360 SELAAEVTIYPQKLVNIRVENSMKERAMEVPAIANIIAKMEDEMAGNRIILVRPSGTEPL 419  
SELAAE+ +PQ LVN+RV + K + E + +I+++E EM G+GRILVRPSGTEPL  
Sbjct: 360 SELAAEMQKFPQLLVNVRVTD--KYKVEENKVKAVISEVEKEMNGDGRILVRPSGTEPL 417

40  
Query: 420 LRVMAEAPTDAEVDYVVDTIADVVRTEIGCD 450  
+RVMAEA T D YV+ I +VVR+E+G +  
Sbjct: 418 VRVMAEAKTKELCDEYVNRIVEVVRSEMGLE 448

An alignment of the GAS and GBS proteins is shown below.

Identities = 400/450 (88%), Positives = 429/450 (94%)

40  
Query: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETDRPRVFVARDTRISGEMLESA 60  
MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHET+RP+VFVARDTRISGEMLESA  
Sbjct: 1 MGKYFGTDGVRGEANVELTPELAFKLGFRGGYVLSQHETERPKVVFVARDTRISGEMLESA 60

45  
Query: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGSDGFKL 120  
LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFG+DGFKL  
Sbjct: 61 LIAGLLSVGIEVYKLGVLATPGVSYLVRTKASAGVMISASHNPALDNGIKFFGNDGFKL 120

50  
Query: 121 DDDRELEIEALLDAKEDTLPRPSAAGGLTLVDYPEGLRKYEFMESTGIDLEGMKVALDT 180  
DD+ELEIEALLDA EDTLPRPSA+GLGTLVDYPEGLRKYEF+ +TG DL GM VALDT  
Sbjct: 121 ADDQELEIEALLDAPEDTLPRPSAEGGLTLVDYPEGLRKYEFVLTGTDLSCMTVALDT 180

55  
Query: 181 ANGAATASARNIFLDLNAIDISVIGDQPDGLNINDGVGSTRPEQLQSLVRENGSDIGLAFD 240  
ANGAA+ SAR++FLDLNA+I+VIG++P+GLNINDGVGST PEQLQ LV+E G+D+GLAFD  
Sbjct: 181 ANGAASVSARDVFLDLNABIAVIGKEKPNGLNINDGVGSTRPEQLQELVKETGADLGLAFD 240

60  
Query: 241 GDSDRLIAVDENGEIVDGDKIMFIIGKYLSDKGQLAQTIVTTVMSNLGFHKALDREGIH 300  
GDSDRLIAVDE GEIVDGD+IMFIIGKYL+KG LA NTIVTTVMSNLGFHKALD++GI+  
Sbjct: 241 GDSDRLIAVDDEGEIVDGDRIIMFIIGKYLSEKGLLAHNTIVTTVMSNLGFHKALDKQGIN 300

65  
Query: 301 KAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLTKVMKETGKKLS 360  
KAITAVGDRYVVEEMR SGYNLGGEQSGHVIIMDYNTTGDGQLTAIQL KVMKETGK LS  
Sbjct: 301 KAITAVGDRYVVEEMRSGYNLGGEQSGHVIIMDYNTTGDGQLTAIQLAKVMKETGKSL 360

70  
Query: 361 ELASEVTIYPQKLVNIRVENNMKDKAMEVPAIAEIIAKMEEEMDGNRIILVRPSGTEPLL 420  
ELA+EVTTIYPQKLVNIRVEN+MK++AMEVPAIA IIAKME+EM GNRIILVRPSGTEPLL

-1516-

Sbjct: 361 ELAAEVTTIYPQKLVNIRVENSMKERAMEVPAIANIIAKMEDEMAGNGRILVVRPSGTEPLL 420

Query: 421 RVMAEAPTNEAVDYVVDTIADVVRTEIGLD 450

RVMAEAPT+ VDYVVDTIADVVRTEIG D

Sbjct: 421 RVMAEAPDAEVDYVVDTIADVVRTEIGCD 450

SEQ ID 4236 (GBS402) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 5; MW 78kDa).

GBS402-GST was purified as shown in Figure 218, lane 3-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1383

A DNA sequence (GBSx1468) was identified in *S.agalactiae* <SEQ ID 4239> which encodes the amino acid sequence <SEQ ID 4240>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]

Identities = 90/324 (27%), Positives = 167/324 (50%), Gaps = 18/324 (5%)

Query: 1 MKKFFTNKFWLGVVSLFLAILLFLTATATSMNHQDNSKIAG-----ASETYTHLTLDVPI 55  
M KF N++ + +++L A+LL++ A + N K G S T TLTD+P+

Sbjct: 1 MDKFLNNRWAVKIIALLFALLLYV---AVNSNQAPTPKKPGESFFPTSTTDEATLTDIPV 57

Query: 56 DIKYDSDDYFISGYSGADVYMS-SVNRVKLDSEINEDTRKFKVVADLTNMKPGTHKVP 114  
YD ++Y ++G +V + S + VK + T+ F++ AD+ ++K GTHKV L

Sbjct: 58 KAYYDDENYVVTGVPQTVNVTIKGSTSAVKKARQ---TKNFEIYADMEHLKTGTHKVEL 113

Query: 115 KVVNLPSGVNATVSPTTITVTMGGKKTKKEFPV-YGHVNDKQIKAGYAVDKMSVDVSKVKV 173  
K N+ G+ +++P+ TVT+ ++ TK FPV + N ++K GY+ ++ V V++

Sbjct: 114 KAKNVSDGLTISINPSVTTVTIQTERTTKSFPVEVEYYNKSMMKKGYSPEQFIVSPKNVQI 173

Query: 174 TSDESIIDRIDHVAANIPDDKVLDDDFNKTVTLQAVTADGTVLASIIHPSKATLSVKVK 233  
T +++ID I A++ + D+ K + DG L + PS ++V V

Sbjct: 174 TGSKNVIDNISLHKASVNLNA-DETIEKEAKVTVDKDGNALPVDVEPSVIKITVPVTS 232

Query: 234 LTKTVPINLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVIN-AEVDISDVTKNT 292  
+K VP + G D +S N + S + + G+++ L+++ I+ +D+S + K++

Sbjct: 233 PSKKVPFKIERTGSLPDGVSIANIESSPSEVTVYGSQDVLDSLEFIDGVSLDLSKINKDS 292

Query: 293 --EKKINLSANNVSDPAQVTVQL 314  
E I L + P++VT+ +

Sbjct: 293 DIEADIPLPDGVKKISPSKVTLHI 316

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4241> which encodes the amino acid sequence <SEQ ID 4242>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-1517-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

# 5 The protein has homology with the following sequences in the databases:

>GP:CAB11969 GB:Z99105 ybbR [Bacillus subtilis]  
 Identities = 81/322 (25%), Positives = 154/322 (47%), Gaps = 15/322 (4%)

Query: 1 MKRFLNSRPWLGMVSVFFAILLFLTAASSNH---NNSSSQIYSPIETYTHSLKDVPI DM 56  
 M +FLN+R + ++++ FA+LL++ A +SN + T +L D+P+  
 Sbjct: 1 MDKFLNNRWAVKIIALLFALLLYV-AVNSNQAPT PKKPGESFFPTSTTDEATLTIDIPVKA 59

Query: 57 KYDS DKYFISGYSYGA EYVLT-STNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNL RV 115  
 YD + Y ++G V + ST+ +K + T+NF+I AD+ H GT V L+  
 Sbjct: 60 YYDDENYVVTGVPQT VNTIKGSTSAVKKARQ----TKNFEIYADMEHLKTGTHKVELKA 115

Query: 116 ENLPSGVTATVSPDKISVTIGKKESKVFVRGS-VDKQIANGYEISKIETGVNKVEVTS 174  
 +N+ G+T +++P +VTI ++ +K FPV + ++ GY + V++T  
 Sbjct: 116 KNVSDGLTISINPSVTITVIQERTTKSFPEVEY YNKSMMKKGYSPQPIVSPKNVQITG 175

Query: 175 DESTIALIDHVVAKLPDDQVLDNRNYSRVTLQAVSADGTILASAIIDPAKTNLSVAVKKIT 234  
 ++ I I A + + D + DG L ++P+ ++V V +  
 Sbjct: 176 SKNVIDNISLHKASVNLENA-DETIEKEAKVTVYDKGNALPVDVEPSVIKITVPTSPS 234

Query: 235 KSVPIRVEAVGMMDDSLSDIQYKLSKQTAVISGSREVLEDIDEII-AE VNISDVTKNT-- 291  
 K VP ++E G + D +S + S + GS++VL+ ++ I +++S + K++  
 Sbjct: 235 KKVPFKIERTGSLPDGVSIA NIESSPSEVT VYGSQDVLDSEFIDGVSLDLSKINKDSI 294

Query: 292 SKTVSLSSSQVSI EPSVVTVQL 313  
 + L I PS VT+ +  
 Sbjct: 295 EADIPLPDGVKKISPSKVTLHI 316

An alignment of the GAS and GBS proteins is shown below.

Identities = 198/319 (62%), Positives = 251/319 (78%), Gaps = 1/319 (0%)

Query: 1 MKKFFFTNKFVLGVVSLFLAILLFLTATATSMNHQDNSKIAGASETYTHLTLDVPIDIKYD 60  
 MK+F ++ WL G+VS+F AILLFLTA A+S ++ +S+I ETYTH+L DVPID+KYD  
 Sbjct: 1 MKRFLNSRPWLGMVSVFFAILLFLTA-ASSNHNSSSQIYSPIETYTHSLKDVPI DMKYD 59

Query: 61 SDDYFISGYSYGADVYMSSVNRVKLDSEINEDTRKFKVADLTFNMKPGTHKVP LKVNL P 120  
 SD YFISGYSYGA+VY++S NR+KLDSE+N DTR FK+VADLT+ PGT V L+V NLP  
 Sbjct: 60 SDKYFISGYSYGA EYVLTSTNRIKLDSEVNNDTRNFKIVADLTHSHPGTVSVNL RVENLP 119

Query: 121 SGVNATVSPTTITVTMGKKKTKEFPVYGHVNDKQIKAGYAVDKMSVDVSKVTSDESII 180  
 SGV ATVSP I+VT+GKK++K FPV G V+ KQI GY + K+ V+KV+VTSDES I  
 Sbjct: 120 SGVTATVSPDKISVTIGKKESKVFVRGSVDKQIANGYEISKIETGVNKVEVTSDESTI 179

Query: 181 DRIDHVAANIPDDKVLDDDFNKTVTTLQAVTADGTVLASIIHPSKATLSVKVKKLT KTVP I 240  
 IDHV A +PDD+VLD +++ VTLQAV+ADGT+LAS I P+K LSV VKK+TK+VPI  
 Sbjct: 180 ALIDHVVAKLPDDQVLDNRNYSRVTLQAVSADGTILASAIIDPAKTNLSVAVKKITKSVPI 239

Query: 241 NLIPVGQFSDSISKINYKLSQEKAVISGTKEALEAISVINA EVDISDVTKNTEKKINLSA 300  
 + VG DS+S I YKLS++ AVISG++E LE I I AEV+ISDVTKNT K ++LS+  
 Sbjct: 240 RVEAVGMMDDSLSDIQYKLSKQTAVISGSREVLEDIDEIIAEVNISDVTKNTSKTVSLSS 299

Query: 301 NNVSVDPAQVTVQLTTT KK 319  
 + VS++P+ VTVQLTTT KK  
 Sbjct: 300 SQVSI EPSVVTVQLTTT KK 318

60 SEQ ID 4240 (GBS99) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 6; MW 35.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 9; MW 60.7kDa).

-1518-

The GBS99-GST fusion product was purified (Figure 197, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 293), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1384

A DNA sequence (GBSx1469) was identified in *S.agalactiae* <SEQ ID 4243> which encodes the amino acid sequence <SEQ ID 4244>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0503(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1385

A DNA sequence (GBSx1470) was identified in *S.agalactiae* <SEQ ID 4245> which encodes the amino acid sequence <SEQ ID 4246>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -9.50    Transmembrane    20 - 36 ( 18 - 46)
      INTEGRAL    Likelihood = -7.64    Transmembrane    48 - 64 ( 42 - 68)
      INTEGRAL    Likelihood = -3.40    Transmembrane    80 - 96 ( 80 - 96)

----- Final Results -----
      bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB11968 GB:Z99105 alternate gene name: ybbQ-similar to
      hypothetical proteins [Bacillus subtilis]
      Identities = 125/253 (49%), Positives = 186/253 (73%), Gaps = 5/253 (1%)

Query: 27 MDIIIVAVLIYKFIKALAGTKIMSLIQGVILFIIIRFVSEWIGLTTITFLMNQIVTYGVI 86
      +DI++V +IYK I + GTK + L++G+++ ++R S+++GL+T+ +LM+Q +T+G +
Sbjct: 16 VDILLVWVVIYKLMIVIRGTVKAVQLLKGIVVIVLVRMASQYLGSLTQLWLMDAQATTWGF 75

Query: 87 AGVVIFAPEIRTGLEKFGRTPLFTQRSQSSDE---KLVDALVKAVAYMSPRKIGALIS 143
      A ++IF PE+R LE+ GR F RS +E K ++A+ KA+ YM+ R+IGAL++
Sbjct: 76 AIIIIFQPELRRALQLGRGR--FFSRSGTPVEEAQQKTIEAITKAINYMAKRRIGALLT 133

Query: 144 IERTQTQLQEYIATGIPLDADISSELLINIFIPNTPPLHDGAVIVKDKKIATACSYLPLSES 203
      IER + +YI TGIPL+A +SELLINIFIPNTPPLHDGAVI+K+ +IA A YLPLSES
Sbjct: 134 IERDTGMGDYIETGIPLNAKVSELLINIFIPNTPPLHDGAVIMKNNIEAAAACYLPLSES 193

Query: 204 SSISKEFGTRHRAAIGLSENSDALTVIVSEETGGISVALKGEFLHDLKDSFEAILRTQL 263
```

-1519-

ISKE GTRHRAA+G+SE +D+LT+IVSEETGG+SVA G+ +L+++ + +L +  
 Sbjct: 194 PFISKELGTRHRAAVGISEVTDLSLTIIVSEETGGVSAKNGDLHRELTEEALKEMLEAEF 253

Query: 264 IQNQEENSKLAWY 276

+N + S WY

Sbjct: 254 KKNTRDTSSNRWY 266

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4247> which encodes the amino acid sequence <SEQ ID 4248>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -6.64 Transmembrane 20 - 36 ( 19 - 40)  
 INTEGRAL Likelihood = -6.21 Transmembrane 48 - 64 ( 47 - 68)  
 INTEGRAL Likelihood = -2.07 Transmembrane 76 - 92 ( 76 - 92)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.3654(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB03984 GB:AP001507 unknown conserved protein [Bacillus halodurans]  
 Identities = 117/255 (45%), Positives = 178/255 (68%), Gaps = 6/255 (2%)  
 Query: 19 PWL-LAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFLVLRFIAEWIGFTTTITYLM 77  
 PWL +LDIL+V Y+IY+ I + GT+ + L++G+ L++ I+ + T+ +++  
 Sbjct: 8 PWLNYLTQILDILVVITYVIYKAIMIIRGTRAVQLKGITVILIVYALSIFFNLRTLGWIV 67  
 Query: 78 NQVITYGVIAGVVIFTPEIRAGLEKFGTRSTQVFLQKQYVSSSEAL---VDALIKSVAYMG 134  
 NQ ITYG++A ++IF PE+R LE+ GR F + + E + +DA++K+ YMG  
 Sbjct: 68 NQAITYGLLAVIIIFQPELRRALEQLGRGR--FFASRTANEETMKKTIDAIVKASTYMG 125  
 Query: 135 PRKIGALIAIEQTQTLOEYIATGIPLNADISSQLLINIFIPNTPLHDGAVIVGQNKIVAA 194  
 R+IGALI++E+ + +Y+ TGIP+NA+++S+LLIN FIPNTPLHDGAVI+ + I+AA  
 Sbjct: 126 KRRIGALISMERETGMTDYVETGIPMNANLTSELLINTFIPNTPLHDGAVIINNDTILAA 185  
 Query: 195 CAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIIVSEETGAISVTRKGQFLHDLSTDE 254  
 YLPLSE+ ISKE GTRHRAA+G+SE +D LTI+VSEETG IS+T+ G+ DL ++  
 Sbjct: 186 ACYLPLSENPFIKELGTRHRAALGVSEVTDCLTIVVSEETGHISLTKNGELHRDLDEEQ 245  
 Query: 255 FETVLRTYLMSNSNV 269  
 ++L L+S + +  
 Sbjct: 246 LRSLEAEELISEAKM 260

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/283 (71%), Positives = 239/283 (84%), Gaps = 2/283 (0%)  
 Query: 1 MDIFSAIDSKFWASIMENPWWILIHLMDIIVAVLIYKFIKALAGTKIMSLIQGVILFII 60  
 M+ S+ID KF S+ +PW++ +HL+DI+IVA LIY+FIKAL GTKIMSL+QGVI F++  
 Sbjct: 1 MNNLSSIDIKFLLSLFADPWLLAVHLLDILIVAYLIYRFIKALTGTKIMSLVQGVIFLVLV 60  
 Query: 61 IRFVSEWIGLTTITFLMNQIVTYGVIAGVVIFAPEIRTGLEKFGTRTPQLFTQRSQSSDE 120  
 +RF++EWIG TTIT+LMNQ++TYGVIAGVVIF PEIR GLEKFG+ Q+F Q+ +SS+  
 Sbjct: 61 LRFIAEWIGFTTTITYLMNQVITYGVIAGVVIFTPEIRAGLEKFGTRSTQVFLQKQYVSS 120  
 Query: 121 KLVDALVKAVAYMSPRKIGALISIERQTLOEYIATGIPLDADISSELLINIFIPNTPLH 180  
 LVDAL+K+VAYM PRKIGALI+IE+TQTLOEYIATGIPL+ADISS+LLINIFIPNTPLH  
 Sbjct: 121 ALVDALIKSVAYMGP RKIGALIAIEQTQTLOEYIATGIPLNADISSQLLINIFIPNTPLH 180  
 Query: 181 DGAVIVKDKKIATACSYLPLSESSSISKEFGTRHRAAIGLSENSDALTIVVSEETGGISV 240  
 DGAVIV KI AC+YLPLSES +ISKEFGTRHRAAIGLSENSDALT+IVSEETG ISV  
 Sbjct: 181 DGAVIVGQNKIVAACAYLPLSESKAISKEFGTRHRAAIGLSENSDALTIIIVSEETGAISV 240  
 Query: 241 ALKGFEFLHDLKDSFEAILRTQLIQNQEENSKLAWYNQLLRK 283

-1520-

KG+FLHDL S D FE +LRT L+ N N L WY ++L K  
 Sbjct: 241 TRKGQFLHDLSTDEFETVLRTYIMSN--SNVTLPWYKKILGGK 281

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1386

A DNA sequence (GBSx1471) was identified in *S.agalactiae* <SEQ ID 4249> which encodes the amino acid sequence <SEQ ID 4250>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.60 Transmembrane 33 - 49 ( 33 - 49)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1387

A DNA sequence (GBSx1472) was identified in *S.agalactiae* <SEQ ID 4251> which encodes the amino acid sequence <SEQ ID 4252>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1001(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9781> which encodes amino acid sequence <SEQ ID 9782> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC84012 GB:AF080002 UDP-N-acetylmuramyl tripeptide synthetase  
 MurC [Heliobacillus mobilis]  
 Identities = 143/442 (32%), Positives = 229/442 (51%), Gaps = 17/442 (3%)  
 Query: 12 GKSAHYLLSKMGRGST-YPGSLALKFKDKILDITIAKDYE--IVVVTGTINGKTLTTALTVG 68  
 GK+A +L + G G T +PG + + IL +A+ + +VVTGTINGKT T+ +  
 Sbjct: 2 GKTAIWLNNRRFGHGGTSFPGGIGRRVAPQILTALARQLKRGAMVVVTGTINGKTITTSKMLAA 61  
 Query: 69 ILKEAFGQVVTPNSGANMITGIVSTFLTAKKSKSG--KKIAVLEIDEASLPRTQYIKPS 126  
 I++++ + N +GAN++ GI + F+ + + ++E+DEA++P++ + ++P  
 Sbjct: 62 IVEKSSSLTLTHNRAGANLVGGITTAFFIDSATIGGSITSDLGIIIEVDEATIPQLVREVQPK 121  
 Query: 127 LFVFTNIFRDQMDRYGEIYTTYQMILDGAANAP-QATILANGDSPLFNS--KSVTPNPVQF 183  
 V TN FRDQ+DR+GE+ T ++ + P Q+ + N D PL S K V +  
 Sbjct: 122 GVVVTNFFRDQLDRFGELDKTVSLVGEALRLLPVQSIIVLNADDPLVASLGKDFPGRVLY 181  
 Query: 184 YGFNTDKHEPRLAHYNTGILCPKCQAILTYRLNTYANLGDYTCPNCDFERPNDLYALTR 243  
 +G + + R + E C C LTY + LG Y C +C FERP +T

-1521-

Sbjct: 182 FGIDDRSYGAREMLQSAETRFRCRLCGHPLTYDWFFGQLGHYRCSHCGERPEPKIKVTG 241

Query: 244 LTHLTNTSSGFVIDGQ---QYNINVGGLYNIYNALAAVSAEYFGVEPSQIKDGFDKSR 299  
 + S F ++ Q ++ G YNIYNALAA++ A + I+ G R

5 Sbjct: 242 IQLKGEESAFETVETPRGTWQLELSTPGFYNIYNALAAIASAIRLDLPEKAIRAGLQGYR 301

Query: 300 AVFGRQETFTIGN-KKCTLVLIKNPVGASQALDMIKLAPYFSLSVLLNANYADGIDTSW 358  
 FGR E + + ++ L LIKNP G + + + P L V++N N ADG D SW

10 Sbjct: 302 TNFGRMERIELEDGRRAFLALIKNPTGCDEVIRTLVQNRGPKRLLVIINDNADGRDISW 361

Query: 359 IWDANFETI--LTMNIPEIFAGGVRHSEIARRLRVTGYDEKRIK-QADKLQDIMTMIEQQ 415  
 +WDA+FE++ + + +F G+R ++A RL TG + I+ +A+ I + +E

Sbjct: 362 LWDADFESELPVPELRSVFTSGLRGEDMALRLNYTGIPAESIRYEANVESAIRSALEMT 421

15 Query: 416 ET-EHAYILATYTAMLEFREIL 436  
 E E YIL TYTA+LE + L

Sbjct: 422 EPGETLYILPTYTALLESKAAL 443

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4253> which encodes the amino acid  
 20 sequence <SEQ ID 4254>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

25 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 343/446 (76%), Positives = 393/446 (87%)

Query: 1 MKINTALGVAAGKSAHYLLSKMGRGSTYPGSLALKFDDKIDLTIAKDYEIVVVTGTNGKT 60  
 MK+ T LG+ AGK+A +L+K+GRGSTYPG LAL DKDIL ++KDY+IVVVTGTNGKT

35 Sbjct: 1 MKMKTLGLIAGKAAQSILTKLGRGSTYPGRLALACDKDILKDLSDYDIVVVTGTNGKT 60

Query: 61 LTTALTVGILKEAFGQVVTNPSGANMITGIVSTFLTAKKSKSGKKIIVLEIDEASLPRIT 120  
 LTTALTVGILKEAFG+++TNPSGANMITGI STFL AKK KS ++IAVLEIDEASLPRIT

Sbjct: 61 LTTALTVGILKEAFGEIITNPSGANMITGITSTFLAAKKGKSERQIAVLEIDEASLPRIT 120

40 Query: 121 QYIKPSLFFVTNIFRDQMDRYGEIYTTYQMILDGAANAPQATILANGDSPLFNSKSVTNP 180  
 Y+KPSLFF+TNIFRDQMDRYGEIYTTYQMI+DGA NAP+ATILANGDSP+F+SK + NP

Sbjct: 121 TYLKPSLFFVTNIFRDQMDRYGEIYTTYQMIVDGARNAPKATILANGDSPIFSSKDIVNP 180

45 Query: 181 VQYGFNTDKHEPRLAHYNTEGILCPKCQAILTYRLNTYANLGDYTCPNCDFERPNLDYA 240  
 VQ+YGF+T KH P+LAHYNTEGILCPKC+ IL YRLNTYANLGD+ C NC F+RP LDY

Sbjct: 181 VQYYGFD TAKHAPQLAHYNTEGILCPKCEHILQYRLNTYANLGDVCLNCQFQRPTLDYQ 240

Query: 241 LTRLTHLTNTSSGFVIDGQYNNINVGGLYNIYNALAAVSAEYFGVEPSQIKDGFDKSRA 300  
 LT LT +T+ SS FVIDGQ Y INVGGLYNIYNALAAVSAE+FGV P +IK GF+KS+A

50 Sbjct: 241 LTELTAITHQSSEFVIDGQYNIKNINVGGLYNIYNALAAVSAEFFGVSPKIKAGFNKSKA 300

Query: 301 VFGRQETFTIGNKKCTLVLIKNPVGASQALDMIKLAPYFSLSVLLNANYADGIDTSWIW 360  
 VFGRQETFT+G+K CTL+LIKNPVGASQAL+MI+LA YPFSLSVLLNANYADGIDTSWIW

55 Sbjct: 301 VFGRQETFTVGDKSCTLILIKNPVGASQALEMIQLADYPFSLSVLLNANYADGIDTSWIW 360

Query: 361 DANFETILTNIPEIFAGGVRHSEIARRLRVTGYDEKRIKQADKLQDIMTMIEQQETEHA 420  
 DANFE I M I EI AGGVRHSEIARRLRVTG+D+ +IKQA+KL+ I+ IE+QE +HA

Sbjct: 361 DANFELITQMPITEINAGGVRHSEIARRLRVTGFDITIKQAEKLEQIETIEKQEAHA 420

60 Query: 421 YILATYTAMLEFREILANHNNAIRKEM 446  
 YILATYTAMLEFR +LA+ + + KEM

Sbjct: 421 YILATYTAMLEFRSLADRHHVVEKEM 446

-1522-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1388

A DNA sequence (GBSx1473) was identified in *S.agalactiae* <SEQ ID 4255> which encodes the amino acid sequence <SEQ ID 4256>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3010(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC84011 GB:AF080002 cobyric acid synthase CobQ [Helicobacillus  
mobilis]  
Identities = 89/250 (35%), Positives = 129/250 (51%), Gaps = 9/250 (3%)

20 Query: 11 TKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCMTFDIVSLED RFPNYYQMA 70  
+K TL + HLY +LLN YGD GNI+ ++ E G + SL ++ + +  
Sbjct: 2 SKTSNRITLHLYPDLLNLYGDRGNIITLRRRCWEWGITLQVHSASLGEKAAFDADLV 61

25 Query: 71 FFGGGQDYEQAIVARDLPSKKEDINKFIQNGV-VLAICGGFQLLGQYYIQANGERIEGI 129  
F GGG D EQ ++ +D K G+ +L++CGG+QLLG YY GE + G+  
Sbjct: 62 FMGGGS DREQTL LFQDFQKHGEPALVEAAEGGLPLL SVCGGYQLLGLYYRTH TGEEMPGL 121

30 Query: 130 GVMGHYTLNQNNNRYIGDIKIHNDEFNE--TYYG FENHQGR TFLSEDE--KPLGTVIYGN 185  
G+ +T + R IG++ E T GFENH GR TFL +PL V G  
Sbjct: 122 GLFDAWT-EAGSTR LIGNVVAQAPLLGEQATLVGFENHSGRTFLGSRGGIQLAQTAGF 180

35 Query: 186 GNNKEDGTGEGVHYKNVFGSYFHPILSRNANLAYRLVATALRNKYG---KEIVLPSYEEI 242  
GNN +D EG YKN G+Y HGP+L +N LA L++ AL +YG + ++E  
Sbjct: 181 GNNGDDQGE GAVYKNAVGTYLHGPVLPKNPALADWLLSKALERRYGGGSLSTLQDTWENR 240

Query: 243 LSLEIPEEYG 252  
L + + +G  
Sbjct: 241 AHL SVAQRFG 250

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4257> which encodes the amino acid sequence <SEQ ID 4258>. Analysis of this protein sequence reveals the following:

Possible site: 32  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.2586(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 197/260 (75%), Positives = 224/260 (85%)

Query: 1 MTTYSLKSPPTKDYKYTLNVAHLYGNLLNTYGDNGNILMMKYVGEKLGCMTFDIVSLED 60  
MTYTS LKSP +DY Y L +AHLYGNL+NTYGDNGNILM+KYV EKL G ++T DIVS+ D  
Sbjct: 1 MTTYSLKSPENQDYIYDLTIAHLYGNLMNTYGDNGNILMLKYVAEKL GARVTVDIVSIND 60

55 Query: 61 RFPNYYQMAFFGGGQDYEQAIVARDLPSKKEDINKFIQNGVVLAI CCGFQLLGQYYIQ 120  
F+ + Y + FFGGGQDYEQ+IVA+DLPSKK + +I NN VVLAICGGFQLLGQYY+Q  
Sbjct: 61 TFEQDDYDIVFFGGGQDYEQSI VAKDLPSKKAALADYIANNKVVLAI CCGFQLLGQYYVQ 120

-1523-

Query: 121 ANGERIEGIGVMGHYTLNQNNNRVIGDIKIHNDEFNETYYGFENHQGRIFLSEDEKPLGT 180  
 ANG +I+G+G+MGHYTLNQ+ NR+IGDIKIHNDEFNETYYGFENHQGRIFLS DEKPLG  
 Sbjct: 121 ANGKIDGLGIMGHYTLNQHQNRFIGDIKIHNDEFNETYYGFENHQGRIFLSGDEKPLGR 180

5 Query: 181 VIYGNNGNKEDGTEGVHYKNVFGSYFHGPILSRNANLAYRLVATALRNKYGKEIVLPSYE 240  
 V+YGNNGNKED TEGVHYKNV+GSYFHGPILSRN NLAYRLV TAL+ KYG I LPSY+  
 Sbjct: 181 VVYGNNGNKEDQTEGVHYKNVYGSYFHGPILSRNVNLAYRLVTTALKKKYGSALSLSYD 240

10 Query: 241 EILSLEIPEEYGDVKSADF 260  
 +IL EI EEY D+KSKA F  
 Sbjct: 241 DILKQEITEEYADLKSASF 260

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 15 Example 1389

A DNA sequence (GBSx1474) was identified in *S.agalactiae* <SEQ ID 4259> which encodes the amino acid sequence <SEQ ID 4260>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1701(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04402 GB:AP001509 lipocate-protein ligase [Bacillus halodurans]  
 Identities = 153/316 (48%), Positives = 212/316 (66%), Gaps = 3/316 (0%)

30 Query: 10 DPAYNVALEAYAFQKLTIDEIFIL-WINEPAIIIGRHQNTIQEINKEFIDKNGIHVVR 68  
 DP N+A+E YA + L DI+E ++L +INEP+IIIGR+QNTI+EIN E+++ NGIHVVR  
 Sbjct: 11 DPRINLAIEEYALKNL-DINETYLLFYINEPSIIIGRNQNTIEEINTEYVESNGIHVVR 69

35 Query: 69 LSGGGAVYHDLNNLNYTIIISNNTQEGAFDFQTFSPVIDTLAKLGVKAEFTGRNDL-EIN 127  
 LSGGGAVYHD NLN++ I+ + E +FQ F+ PVI LAKLGV AE GRND+ +  
 Sbjct: 70 LSGGGAVYHDHGNLNFSTFKDGSFNFQKFTDPVIKALAKLGVTAELKGRNDIIASD 129

40 Query: 128 GQKFAGNAQAYYKGRMMHHCCLLFDVDMSVLGQALKVSKDKIESKGIKSVRARVTNIVDH 187  
 G+K +GNAQ KGRM HG LLFD ++ + AL VSKDKIESKGIKS+R+RV NI +  
 Sbjct: 130 GRKISGNAQFSTKGRMFHSGTLLFDEIDHVVSAALNVSKDKIESKGIKSIRSRVANISEF 189

45 Query: 188 LSDKITVQEFSDAILAQMKEEYPEMDEYVLSDAELSEIQAMRDNQFATWDWTYGKAPEYT 247  
 L++KI++ +F +L + + + EY L+ + +EI + ++ WDW YGK+P +  
 Sbjct: 190 LTEKISIDQFRSLLESIFDGQANIQYKLTADDWAEIHLSKERYQNWDWNYGKSPAFN 249

50 Query: 248 IERGVRYPAKITTYANVENSTIKSVKIFGDFGVPVDDIEKMLEGVRYDYKDVLAAALK 307  
 ++ R+P G I V+ TI+ KIFGDFG V D+E L G+RY+ D+ AL  
 Sbjct: 250 LQSHSRFPVGNIDIRLEVKGGTIQCKIFGDFGTDVDRDLEDRLVGIRYERADIEQALA 309

55 Query: 308 TVDTSQYFSRMTPEEI 323  
 VD YF ++ ++I  
 Sbjct: 310 DVDVKTYFGQVEKDDI 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4261> which encodes the amino acid sequence <SEQ ID 4262>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1271(Affirmative) < succ>

-1524-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5       Identities = 249/328 (75%), Positives = 292/328 (88%)

Query: 1   MKYIVNTSNDPAYNVALEAYAFQKLT DIDEIFILWINEPAIIIGRHQNTIQEINKEFIDK 60  
           MKYIVN S++PA+N+ALEAYAF++L + DE+FILWINEPAIIIG+HQNTIQEINKE+ID+  
 10   Sbjct: 1   MKYIVN KSHNPAPNTALEAYAFRELVEEDEL FILWINEPAIIIGKHQNTIQEINKEYIDE 60

Query: 61   NGIHVVRRLSGGGAVYHDLNNLNNTIISNNTQEGAFDFQTFSPVIDTLAKLGKAEFTG 120  
           +GIHVVRRLSGGGAVYHDLNNLNNTIISN T EGAFDF+TFS+PVI TLA LGV A FTG  
 15   Sbjct: 61   HGIHVVRRLSGGGAVYHDLNNLNNTIISNKTAEAGAFDFKTFSPVIATLADLGVTANFTG 120

Query: 121   RNDLEINGQKFAGNAQAYYKGRMMHHGCLLFDVDM SVLGQALKVSKDKIESKGIKSVRAR 180  
           RND+EI+G+K   GNAQAYYKGRMMHHGCLLFDVDM+VLG ALK VSKDKIESKG+KSVRAR  
 20   Sbjct: 121   RNDIEIDGKKICGNAQAYYKGRMMHHGCLLFDVDMTVLGDALKVSKDKIESKGVKSVRAR 180

Query: 181   VTNIVDHLSDKITVQEFSDAILAQMKEEYPEMDEYVLSDAELSEIQAMRDNQFATWDWTY 240  
           VTNI++ L +KITV+EFSD ILA+MKE YP+M EYVLS+ EL++I+   .QF +WDWTY  
 25   Sbjct: 181   VTNINELPEKITVEEFSDKILAKMKETY PDMTEYVLSDELAKIEQSAKEQFGSWDWTY 240

Query: 241   GKAPEYTIERGVRYPAGKITTYANVENSTIKSVKIFGDFFGVKPVDDIEKMLEGVRYDYK 300  
           GKAPEYTIER VRYPAGKI+T+ANVENS IK++KI+GDFFG+K V DIE +L G +Y+Y+  
 30   Sbjct: 241   GKAPEYTIERNVRYPAGKISTFANVENSIIKNLKIYGDFFGIKDVQDIENLLIGCKYEYR 300

Query: 301   DVLAAKLTVDTSQYFSRMTPEEITKAIV 328  
           DV   LKT+DT+QYFSRMT EE+ KAIV  
 35   Sbjct: 301   DVFERLKTIDTTQYFSRMTVEEVAKAIV 328

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1390

35   A DNA sequence (GBSx1475) was identified in *S. agalactiae* <SEQ ID 4263> which encodes the amino acid sequence <SEQ ID 4264>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL   Likelihood = -1.70   Transmembrane   294 - 310 ( 294 - 312)

40   ----- Final Results -----  
           bacterial membrane --- Certainty=0.1680 (Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45   The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA21748 GB:L31844 dihydrolipoamide dehydrogenase [Clostridium  
           magnum]  
       Identities = 229/589 (38%), Positives = 339/589 (56%), Gaps = 25/589 (4%)

50   Query: 1   MAFDVIMPKLGVD MQEGEILEWKKNEGDTVNEGDVLEIMSDKTNMEIEAEDTGVLKIV 60  
           MA V+MPKLG+ M EG ++ WKK EGD V G++L E+ +DK E+E+ D G++ K++  
       Sbjct: 1   MAKIVVMPKLGITMTTEGTLVTWKKAEQDQKVG EILFEVSTDKLTNEVESSDEGIVRKLL 60

55   Query: 61   HQAGDVVPVTEVIAYIGE EGEVGTSSPSADATITAEDGQSVSGPAAPSQETVAAATPK 120  
           GDVV    +A IG E++ +                    +G S   +A   +T A PK+  
       Sbjct: 61   VNEGDVVECLNPVAIIGSADEDISSLL-----NGSSEGS SAEQSDTKA---PKK 107

60   Query: 121   ELA ADEY--DIVVVG GGPAGYAAIRGAQLGKIAIVEKTEFGGTCLNVGCIPTKTYLKN 178  
           E+ A +   ++VV+GGGP GY AAIR AQLG K+ ++EK GGTCLNVGCIPTK L +  
       Sbjct: 108   EVEAVKGGDNLVVIGGGPGGYVAAIRAAQLGAKVT LIEKESLGGTCLNVGCIPTKVLLHS 167

-1525-

Query: 179 AEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFNGLG 238  
 +++L +K GI++ + ++ K V+K L GV GLL NKV++ G  
 Sbjct: 168 SQLLTEMKEGDKLGIDIEGS-IVVNWKHIQKRKKIVIKKLVSGLLTCKNVKVIKGT 226

5 Query: 239 QVNPDKSVVIGDK-----VIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDREIPK 293  
 + ++++ + + N ++ATGS I G + V+ S L L P+  
 Sbjct: 227 KFESKDTILVTKEGVAEKVNFDNAIATGSMPPFIEGKLSGVIDSTGALSLESNP 286

10 Query: 294 SLAVMGGGVVIGELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKTS 353  
 S+A++GGGV+G+E ++ S G V++IEM I+P MD+E+S + L + G+ I +  
 Sbjct: 287 SIAIIGGGVIGVEFASIFNSLCKVKSIIEMLPFILPPMDREISEIAKAKLIRDGININN 346

15 Query: 354 VGVSEIVEANNQLTLKL--NNGEEVV-ADKALLSIGRVPQMNGLENLEPELEMERGRIKV 410  
 V+ I + + L + + GEE + +K L+++GR + GL+ + ++ E G I V  
 Sbjct: 347 CKVTRIEQGEDGLKVSFIGDKGEESIDVEKVLIAVGRRSNIEGLDVEKIGVKTEGGSII 406

20 Query: 411 NAYQETSIPGIYAPGDVNGTRMLAHAAAYRMGEVAAENALGGNKRKAHLDFTPAAVYTHPE 470  
 N ET++ GIYA GD G MLAH A G VAAEN +G NK K PA VYT PE  
 Sbjct: 407 NDKMETNVGEIYAIGDCTGKIMLAHVASDQGVVAAENIMQNK-KMDYKTVPAVCVYTKPE 465

25 Query: 471 VAMVGMTTEEQAREQYGDILVGKNSFTGNGRAIASNEAHGFVKVIAEPKYKEILGVHIIGP 530  
 +A VG+TEEQAE+ D VGK NG+++ NE G +K+I + KY+ELGVHI+GP  
 Sbjct: 466 LASVGLTEEQAKEKGIDYKVGKFLAANGKSLIMNETGGVIKIITDKKYEELGVHILGP 525

Query: 531 AAAELINEASTIMENELTVYDVAQSIHGHPTFSEVMYEAFLDVLGEAIH 579  
 A +LI EA+ + E T+ ++ ++H HPT E M EA L V +AIH  
 Sbjct: 526 RATDLITEAALALRLEATLEEIITTVHAHPTVGEAMKEAALAVNNQAIH 574

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1819> which encodes the amino acid  
 sequence <SEQ ID 1820>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.70 Transmembrane 297 - 313 ( 297 - 315)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 497/591 (84%), Positives = 538/591 (90%), Gaps = 10/591 (1%)

Query: 1 MAFDVMIPKLGVDMQEIGEILEWKKNEGDTVNEGDVLLLEIMSDKTNMEIEAEDTGVLKIV 60  
 MA ++IMPKLGVDMQEIEI+EWKK EGDTVNEGD+LLEIMSDKTNME+EAED+GVLLKI.  
 Sbjct: 1 MAVEIIMPKLGVDMQEIEI+EWKKQEGDTVNEGDILLEIMSDKTNMELEAEDSGVLLKIT 60

Query: 61 HQAGDVVPVTEVIAYIGEEGEEVGTSSPSA---DATITAEDGQS--VSGPAAPSQETVAA 115  
 QAG+ VPVTEVI YIG EGE V SSP+A + T ED ++ + P AP+Q A+  
 Sbjct: 61 RQAGETVPVTEVIGYIGAEGESVEVSSPAASDVNVARTTEDLEAAGLEVPKAPAQ--AAS 118

Query: 116 ATPKEELAADEYDIVVGGGPAGYAAIRGAQLGGKIAIVEKTEFGGTCLNVGCIPTKTY 175  
 A PK LA DEYDI+VVGGPAGYAAIRGAQLGGKIAIVEK+EFGGTCLNVGCIPTKTY  
 Sbjct: 119 AAPKAALADDEYDIIVVGGGPAGYAAIRGAQLGGKIAIVEKSEFGGTCLNVGCIPTKTY 178

Query: 176 LKNAEILDGLKVAAGRGINLASTNYAIDMDKTVAFKNSVVKTLTGGVRGLLKANKVEIFN 235  
 LKNAEILDG+K+AAGRGINLASTNY IDMDKT V FKN+VVKTLTGGV+GLLKANKV IFN  
 Sbjct: 179 LKNAEILDGIKIAAGRGINLASTNYTIDMDKTVDKNTVVKTLTGGVQGLLKANKVTIFN 238

Query: 236 GLGQVNPDKSVVIGDKVIKGRNVVLATGSKVSRINIPGIESPLVLTSDDILDREIPKSL 295  
 GLGQVNPDK+V IG + IKGRNV+LATGSKVSRINIPGI+S LVLTSDDILDRE+PKSL  
 Sbjct: 239 GLGQVNPDKTVTIGSQTIKGRNVILATGSKVSRINIPGIDSKLVLTSDDILDREMPKSL 298

Query: 296 AVMGGGVVIGELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILAKKGMKIKTSVG 355  
 AVMGGGVVIGELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKIL+KKGKMKIKTSVG  
 Sbjct: 299 AVMGGGVVIGELGLVWASYGVDVTVIEMADRIIPAMDKEVSLELQKILSKKGMKIKTSVG 358

-1526-

Query: 356 VSEIVEANNQLTLKLNNGEEVVDKALLSIGRVPQMNGLENLEPELEMERGRIKVNAVQE 415  
 VSEIVEANNQLTLKLNNGEEVVA+KALLSIGRV QMNGLENL LEM+R RIKVN YQE  
 Sbjct: 359 VSEIVEANNQLTLKLNNGEEVVAEKALLSIGRVSQMNGLENL--NLEMDRNRIKVNDYQE 416

Query: 416 TSIPGIYAPGDVNGTRMLAHAAYRMGEVAAENALGCN-KRKAHLDFTPAAVYTHPEVAMV 474  
 TSIPGIYAPGDVNGT+MLAHAAYRMGEVAAENA+ GN RKA+L +TPAAVYTHPEVAMV  
 Sbjct: 417 TSIPGIYAPGDVNGTKMLAHAAYRMGEVAAENAMHGNTTTRKANLKYTPAAVYTHPEVAMV 476

Query: 475 GMTEEQAREQYGDILVGKNSFTGNGRAIASNEAHGFVKVIAEPKYKEILGVHIIIGPAAAE 534  
 G+TEEQAREQYGD+L+GKNSFTGNGRAIASNEAHGFVKVIA+ KY EILGVHIIIGPAAAE  
 Sbjct: 477 GLTEEQAREQYGDVLIGKNSFTGNGRAIASNEAHGFVKVIADAKYHEILGVHIIIGPAAAE 536

Query: 535 LINEASTIMENELTVYDVAQSIHGHTFSEVMYEAFLDVLGEAIHNPPKRK 585  
 +INEA+TIME+ELTV ++ SIHGHTFSEVMYEAFLDVLGEAIHNPPKRK  
 Sbjct: 537 MINEAATIMSELTVDLLLSIHGHTFSEVMYEAFLDVLGEAIHNPPKRK 587

SEQ ID 4264 (GBS681) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 165 (lane 2; MW 68.3kDa) and in Figure 188 (lane 10; MW 68kDa).

Purified GBS681-His is shown in Figure 240, lane 5-6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1391

A DNA sequence (GBSx1476) was identified in *S.galactiae* <SEQ ID 4265> which encodes the amino acid sequence <SEQ ID 4266>. This protein is predicted to be dihydrolipoamide acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

#### ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4466(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04497 GB:AP001509 dihydrolipoamide S-acetyltransferase  
 [Bacillus halodurans]  
 Identities = 187/462 (40%), Positives = 266/462 (57%), Gaps = 26/462 (5%)

Query: 1 MAVIIMPKLGVDMQEGEILEWKKQVGDVVNEGDVLEIMSDKTNMEIEAEDSGVLLKIT 60  
 MA EI MPKL MQEG +L+W K+ GD V G+ L EIM+DK N+E+EA + G LLK  
 Sbjct: 1 MAKEIFMPKLSSTMQEGTLLQWFKBEGDRVEVGEPLFEIMTDKINIEVEAYEEGTLLKRY 60

Query: 61 HGNGDVVPVTETIGYIGAEGEEVTEASSSENTSVEENATQVTSEPEKVEETSEPSVPAAT 120  
 +G D +PV IGYIG E V +E E T E T+ P++  
 Sbjct: 61 YGEDDEIPVNHVIGYIGTPDESVP----TEPPGASEITASSTDEAGDHRTTAVKKAPSSD 116

Query: 121 SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITPLARRIAEDQ 180  
 E VRATPAAR++A+E IDL V G+G GRV DV FK + TPLA+++AE +  
 Sbjct: 117 R-ENVRATPAARRIAKEKRIDLROVEGSGPEGRVQAVDVATFKKKQKATPLAKKVAEVK 175

Query: 181 GVDIAEITGSGIRGKIVKNDVLAAMSPQAAEAPVETKATPTTEEKQLPEGVEVIKMSAMR 240  
 GV + ++ GSG GK+ + DV A A +PVE K +K+S +R  
 Sbjct: 176 GVALEKVQSGPYGKVYREDVEHAQ----AASPVEDKGNR-----VKLSGLR 218

Query: 241 KAISKGMTNSYLTAPSFTLNVDIDMTMMALRKKLIDPIMAKTGLKVSFTDLIGMAVVKT 300  
 K ++K M +S +AP T+ +IDM+ + +R +L+ I +TG ++S+T+++ AV  
 Sbjct: 219 KVVAKRMVDSAFSAPHVTITTEIDMSSTIKIRSQLGMIEQETGYRLSYTEIVMKAVAHA 278

-1527-

Query: 301 LMKPEHRYLNASLINDAQEIELHNFVNIGIAVGLDDGLIVPVVHNADQMSLSDFVIASKD 360  
 LM H +NAS + EI H V+IG+AV ++ GL+VPVV + D+ L+ K  
 Sbjct: 279 LMS--HPTINASFFEN--EIVYHEDVHIGLAVAVEGGLVVPVVKHVDKGLAQLTNECKT 334

Query: 361 VIKKTQEGKLSAEMSGSTFSITNLGMFGTKTFNPIINQPNNSAILGVGATIPTPTVVDGE 420  
 V ++ +L MSG TF+I+NLGM+ F P+INQP SAILGVG P +DG+  
 Sbjct: 335 VAMAARDNRLSQEMMSGGTFITISNLGMYAIDVFTPVINQPNNSAILGVGRIQEKPVGIDGQ 394

Query: 421 IVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462  
 I RP+M L+ DHR++DG A F+ D+K+++E PF LL+  
 Sbjct: 395 IELRPMMTASLSFDHRVIDGAPAAAFITDVKSMLEQPFQLLM 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4267> which encodes the amino acid sequence <SEQ ID 4268>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4774 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 354/473 (74%), Positives = 390/473 (81%), Gaps = 15/473 (3%)

Query: 1 MAVEIIMPKLGVDMQEGEILEWKKQVGDVNEGDVLEIMSDKTNMEIEAEDSGVLLKIT 60  
 MA EIIMPKLGVDMQEGEI+EWKKQ GD VNEGD+LLEIMSDKTNME+EAEDSGVLLKIT  
 Sbjct: 1 MAFEIIMPKLGVDMQEGEIIWKKQEGDVTNEGDILLEIMSDKTNMELEAEDSGVLLKIT 60

Query: 61 HGNGDVVPVTETIGYIGAEGEVTEASSSENTS-----VEENATQVTSEPEKVEETSEPS 115  
 GD VVPTE IGYIGAEGE V +SSE T+ +A + E V + P  
 Sbjct: 61 RQAGDTPVPTVEVIGYIGAEGESVDTTASSEKTTEIPVPASADAGPAVAPKENVASPA-PQ 119

Query: 116 VPAAT----SGEKVRATPAARKLAREMSIDLALVSGTGANGRVHREDVENFKGAQPRITP 171  
 V A +G KVRATPAARK A EM IDL V GTG GRVH+EDVENFKGAQP+ +P  
 Sbjct: 120 VAATAIPQNGGKVRATPAARKAAEMGIDLGQVPGTGPKGRVHKEDVENFKGAQPKASP 179

Query: 172 LARRIAEQGVDAIEITGSGIRKIVKNDVLAAMSPQAAEAPVETKATPTTEEK--QLPE 229  
 LAR+IA D+G+D+A ++G+G GK++K D++A + A P E KA EEK LPE  
 Sbjct: 180 LARKIAADKGIDLATVSGTGFGKVMKEDIMAILL---AAKPAEAKAPAAKEEKVVDLPE 236

Query: 230 GVEVIKMSAMRKAISKGMTNSYLTAPSFTLNVDIDMTMALRKKLIDPIMAKTGLKVSF 289  
 GVE MSAMRKAISKGMTNSYLTAP+FTLNVDIDMTM+ALRKKLIDPIMAKTGLKVSF  
 Sbjct: 237 GVEHKPMSAMRKAISKGMTNSYLTAPTFTLNVDIDMTMIALRKKLIDPIMAKTGLKVSF 296

Query: 290 TDLIGMAVVKTLMKPEHRYLNASLINDAQEIELHNFVNIGIAVGLDDGLIVPVVHNADQM 349  
 TDLIGMAVVKTLMKPEH Y+NASLINDA +IELH FVN+GIAVGLDDGLIVPV+H A++M  
 Sbjct: 297 TDLIGMAVVKTLMKPEHEYMNASLINDANDIELHRFVNGLGIAVGLDDGLIVPVHGANM 356

Query: 350 LSDFVIASKDVIKKTQEGKLSAEMSGSTFSITNLGMFGTKTFNPIINQPNNSAILGVGA 409  
 LSDFV+ASKDVIK Q GKLL+AEMSGSTFSITNLGMFGTKTFNPIINQPNNSAILGVGA  
 Sbjct: 357 CLSDFVLASKDVIKKAQTGKLLAAEMSGSTFSITNLGMFGTKTFNPIINQPNNSAILGVGA 416

Query: 410 TIPTPTVVDGEIVARPIMAMCLTIDHRIVDGMNGAKFMVDLKNLMENPFGLLI 462  
 TIPTPTVVDGEIV+RPIMAMCLTIDHR+VDGMNGAKFMVDLKNLMENPF LLI  
 Sbjct: 417 TIPTPTVVDGEIVSRPIMAMCLTIDHRLVDGMNGAKFMVDLKKLMENPFELLI 469

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1528-

**Example 1392**

A DNA sequence (GBSx1477) was identified in *S. agalactiae* <SEQ ID 4269> which encodes the amino acid sequence <SEQ ID 4270>. This protein is predicted to be acetoin dehydrogenase (TPP-dependent) beta chain (pdhB). Analysis of this protein sequence reveals the following:

```

5   Possible site: 18
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10   bacterial cytoplasm --- Certainty=0.1267(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9779> which encodes amino acid sequence <SEQ ID 9780> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
    chain [Bacillus halodurans]
    Identities = 189/319 (59%), Positives = 249/319 (77%), Gaps = 1/319 (0%)

20   Query: 11  EAINVAMSEEMRKDEKVFIMGEDVGVYGGDFGTSVGMLEEFGAKRVRDTPISEAAIAGSA 70
    EAI  AM+ EMRK+E VF++GED+GVYGG FG + GM+EEFG++RVR+TPISEAAI+G+A
    Sbjct: 8    EAIREAMTLEMRKNEDVFILGEDIGVYGGAFGVTRGMIEEFGSERVRNTPISEAAISGTA 67

25   Query: 71  IGAAQTGLRPIVDLTFMDFVTIAMDIAVNQGAKTNYMFGGGLSTPVTFRVASGSGIGSAA 130
    IGAA TG+RPI++L F DF+TIAMD +VNQ AK YM+GG P+ R +GSG G+AA
    Sbjct: 68  IGAAITGMRPIELQFSDFITIAMDNMVNQAALRYMYGGKAKVPMVLRTPAGSGTGAAA 127

30   Query: 131 QHSQSLEAWLTHIPGLKVVPAGTVNESKALLKSSILDNNPVIFLEPKALYGKKEEVNMDP 190
    QHSQSLEAW+THIPGLKV P T ++K LLK++I DNNPVIF E K Y K V +
    Sbjct: 128 QHSQSLEAWMTHIPGLKVVPATAYDAKGLLKAIDNNPVIFYEHKLCYRTKCHV-PEE 186

    Query: 191 DFYIPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAEEGINVEVVDPRTLIPLDKELI 250
    ++ IPLGK D+KR+GTD+T+V+ M+ + ++AA E+ +EGI+VEV+DPRTL+PLD+E I
    Sbjct: 187 EYSIPLGKADVKKRGTDVTVVATAMVMHKALEAAVELEKEGISVEVIDPRTLVLDEETI 246

35   Query: 251 IDSVKKTGKLILVNDAYKTGGFTGEIATMVAESEAFDYLDHPVIRLASEDVPVPYSRVLE 310
    I SVKKT +LI+V++A K GGF GEIA+++AESEAFDYLD PI RL + VP+PY+ LE
    Sbjct: 247 IRSVKKTSLRIVVHEAVKRGFGGEIASIIAESEAFDYLDAPIKRLGGKPVPIPNPTLE 306

40   Query: 311 QGILPDVAKIKDAIYKVVN 329
    + +P V I +A+ + +N
    Sbjct: 307 RAAIPQVPDIIEAVKETLN 325

```

45 A related DNA sequence was identified in *S. pyogenes* <SEQ ID 4271> which encodes the amino acid sequence <SEQ ID 4272>. Analysis of this protein sequence reveals the following:

```

    Possible site: 18
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.00    Transmembrane    81 - 97 ( 81 - 97)

50   ----- Final Results -----
    bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

```

>GP:BAB04496 GB:AP001509 acetoin dehydrogenase (TPP-dependent) beta
    chain [Bacillus halodurans]
    Identities = 187/319 (58%), Positives = 244/319 (75%), Gaps = 1/319 (0%)

```

-1529-

Query: 11 EAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTPISEAAISGAA 70  
EA+ AMT EMRK+E++F++GED+GVYGG FG + GMIEEFG +RV++TPISEAAISG A  
Sbjct: 8 EAIREAMTLEMRKNEVDVFLGEDIGVYGGAFGVTRGMIEEFGSEVRNTPISEAAISGTA 67

5 Query: 71 IGAAITGLRPIVDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRVASGSGIGSAA 130  
IGAA+TG+RPI+++ F DF+TI MD +VN AK YM+GG P+ R +GSG G+AA  
Sbjct: 68 IGAAITGLRPIVDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRVASGSGIGSAA 127

10 Query: 131 QHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKS AIRDNNIVLFMEPKALY GKKEEVNQDP 190  
QHSQSLEAW+THIPG+KVV P A DAKGLLK+AI DNN V+F E K Y K V ++  
Sbjct: 128 QHSQSLEAWMTHIPGLKVVQPATAYDAKGLLKAAIDNNPVIFYEHKLCYRTKCHVPEE- 186

15 Query: 191 DFIYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVDPRTLIPLDKELI 250  
++ IPLGK D+KR+GTD+T+V+ M+ + L+AA E+ +GI+VEV+DPRTL+PLD+E I  
Sbjct: 187 EYSIPLGKADVKRKGTDVTVVATAMVHKALEAAVELEKEGISVEVIDPRTLVLDEETI 246

20 Query: 251 IESVKKTKGLMLVNDAYKTGGFGEIATMITESEAFDYLDHPIVRLASEDVPVPYARVLE 310  
I SVKKT +L++V++A K GGF GEIA++I ESEAFDYLD PI RL + VP+PY LE  
Sbjct: 247 IRSVKKTSRLIVVHEAVKRGFGGEIASIIAEESEAFDYLDAPIKRLGGKPVPIPNPTLE 306

Query: 311 QAILPDVEKIKAAIVKMAN 329  
+A +P V I A+ + N  
Sbjct: 307 RAAIPQVPDIIEAVKETLN 325

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 286/331 (86%); Positives = 310/331 (93%)

Query: 1 MSETKVMALREAINVAMSEEMRKDEKVFMLGEDVGVYGGDFGTSVGMLEEFGA KRVRDTP 60  
MSETK+MALREA+N+AM+EEMRKDE +FLMGEDVGVYGGDFGTSVGM+EEFG KRV+DTP  
30 Sbjct: 1 MSETKLMALREAVNLAMTEEMRKDENIFLMGEDVGVYGGDFGTSVGMIEEFGPKRVKDTPI 60

Query: 61 ISEAAIAGSAGAAQTGLRPIVDLTFMDFVTIAMDAIVNQAKTNYMFGGGLSTPVTFRV 120  
ISEAAI+G+AIGAA TGLRPIVD+TFMDF+TI MDAIVN GAK NYMFGGGL TPVTFRV  
35 Sbjct: 61 ISEAAISGAAIGAAITGLRPIVDVTFMDFLTIMMDAIVNNGAKNNYMFGGGLITPVTFRV 120

Query: 121 ASGSGIGSAAQHSQSLEAWLTHIPGLKVVAPGTVNESKALLKSSILDNNPVIFLEPKALY 180  
ASGSGIGSAAQHSQSLEAWLTHIPG+KVVAPG N++K LLKS+I DNN V+F+EPKALY  
40 Sbjct: 121 ASGSGIGSAAQHSQSLEAWLTHIPGIKVVAPGNANDAKGLLKS AIRDNNIVLFMEPKALY 180

Query: 181 GKKEEVNMDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVMQAAEEVAEEGINVEVDP 240  
GKKEEVN DPDFYIPLGKGDIKREGTDLTIVSYGRMLERV+QAAEEVA +GINVEVDP  
45 Sbjct: 181 GKKEEVNQDPDFYIPLGKGDIKREGTDLTIVSYGRMLERVLQAAEEVAADGINVEVDP 240

Query: 241 TLIPLDKELIIDSVKKTGKLILVNDAYKTGGFTGEIATMVAESEAFDYLDHPIVRLASED 300  
TLIPLDKELII+SVKKTGKL+LVNDAYKTGGF GEIATM+ ESEAFDYLDHPIVRLASED  
50 Sbjct: 241 TLIPLDKELIIESVKKTGKLMLVNDAYKTGGFGEIATMITESEAFDYLDHPIVRLASED 300

Query: 301 VFPVYSRVLEQGILPDVAKIKDAIKVNVNKG 331  
VFPVY+RVLEQ ILPDV KIK AI K+ NKG  
Sbjct: 301 VFPVYARVLEQAILPDVEKIKAAIVKMANKG 331

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1393

55 A DNA sequence (GBSx1478) was identified in *S. agalactiae* <SEQ ID 4273> which encodes the amino acid sequence <SEQ ID 4274>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.03 Transmembrane 161 - 177 ( 161 - 178)

----- Final Results -----

bacterial membrane --- Certainty=0.2211(Affirmative) < succ>

-1530-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9777> which encodes amino acid sequence <SEQ ID 9778> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04495 GB:AP001509 acetoin dehydrogenase (TPP-dependent) alpha  
 chain [Bacillus halodurans]  
 Identities = 148/317 (46%), Positives = 214/317 (66%), Gaps = 1/317 (0%)

Query: 8 LSKEQHLDMLKMQRIRDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDLTDSDIIF 67  
 +++++ +D+F +M IR + K ++ +G + G TH +VG+EA++VG+I L + D +  
 Sbjct: 10 MTEKKLVDLFKQMWLIRYFEEKVDEFFAKGMIHGTTHLAVGQEASAVGSIIVLEERDKLT 69

Query: 68 SNHRGHGQTIAGIDIGGMFAELAGKATGTSKGRGGSMLANLEKGNYGTINGIVGGGYAL 127  
 S HRGHG IAKG D+ M AEL G+ TG KG+GGSMH+A++E+GN G NGIVGGG+++  
 Sbjct: 70 STHRGHGHCIAGADVNRMMAELEFGRETGYCKGKGGSMHIAVERGNLGANGIVGGGFSI 129

Query: 128 AVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPIVFFIINNRYGISTDI 187  
 A GAALT + + +V+ F GD A+NEGSFHE+VNLA++W LPV+F NN+YG+S +  
 Sbjct: 130 ATGAALTSKMKKEGYVVLCFFGDGASNEGSFHEAVNLASIWKLPIVVFICENNQYGMSSGV 189

Query: 188 TYSTKIPHYLMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVEVESYRWF 247  
 I H+ RA YGIPG V DGND+ AV + ++ R G GP IVE ++YRW  
 Sbjct: 190 KEMINIEHISDRAAGYGIPG-MVVDGNDVFAVMNVVGRAVDRARRGEGPTIVEAKTYRWK 248

Query: 248 GHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEEELAAIEAQVIKEVEEGVKF 307  
 GHS +DA YRT+EE W+ KDP+ R RA L++ I TEE +I+ + +++E+ V+F  
 Sbjct: 249 GHKSDAKKYRTREEKEWREKDPIARLRATLVKEGIVTEEEADSIQEEAKQKIEDSVQF 308

Query: 308 AEESPPFPMDSVAFEDVF 324  
 A SP P++ EDV+  
 Sbjct: 309 ARNSPEPEIESLLEDVY 325

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4275> which encodes the amino acid sequence <SEQ ID 4276>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3502 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 244/326 (74%), Positives = 278/326 (84%)

Query: 1 MEVMTVLSKEQHLDMLKMQRIRDVDMKFNKLVRRGFVQGMTHFSVGEEAASVGAIQDL 60  
 ME MVT+SKEQHLDMLKMQRIR+ D + NKLVRRGFVQGMTHFSVGEEAA+VGA+ L  
 Sbjct: 1 MEAEMVTVSKEQHLDMLKMERIREFDSRINKLVRRGFVQGMTHFSVGEEAANVGAVHL 60

Query: 61 TDSDIIFSNHRGHGQTIAGIDIGGMFAELAGKATGTSKGRGGSMLANLEKGNYGTINGI 120  
 + DIIFSNHRGHGQ+IAK +D+ M AELAGKATG SKGRGGSMLA+ EKGNYGTNGI  
 Sbjct: 61 SYDDIIFSNHRGHGQSIADMDLNMMAELAGKATGVSKGRGGSMLADFEKGNYGTINGI 120

Query: 121 VGGGYALAVGAALTQQYEGTDNIVIAFSGDSATNEGSFHESVNLAAVWNLPIVFFIINNRR 180  
 VGGGYALAVGAALTQQY+GT+NI +AFSGD ATNEGSFHESVN+AA W LPVFFIINNRR  
 Sbjct: 121 VGGGYALAVGAALTQQYKGTNNIAVAFSGDGATNEGSFHESVNMAATWKLPVFFIINNRR 180

Query: 181 YGISTDITYSTKIPHYLMRADAYGIPGHYVEDGNDLMAVYEKMHEVINYVRSGNGPAIVE 240  
 YGIS I +T PHLY RA+AYG+PG Y EDGND+MAVYE M + + +VR GNGPAIVE  
 Sbjct: 181 YGISMSINNATNTPHLYTRAEAYGVPGFYCEDGNDVMAVYETMGKAVEHVRGNGPAIVE 240

Query: 241 VESYRWFGHSTADAGVYRTKEEVDSWKAKDPVKRYRAYLIENEIATEEELAAIEAQVIKE 300  
VESYRWFGHSTADAG YRTKEEVD WK KDP+ +YR YL IAT++EL AI+AQV KE  
Sbjct: 241 VESYRWFGHSTADAGKYRTKEEVDDEWKEKDPMIKYRTYLTSEGIATDDELDIAIQAVKKE 300

Query: 301 VEEGVKFAEESPPFDMSVAFEDVFVD 326  
V++ +FA+ SP P++SVAFEDV+VD  
Sbjct: 301 VDDAYEFAQNSPDPELSVAFEDVWVD 326

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score: . . -14.75
GvH: Signal Score (-7.5): -4.24
    Possible site: 48
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value: -3.03 threshold: 0.0
    INTEGRAL    Likelihood = -3.03   Transmembrane  161 - 177 ( 161 - 178)
    PERIPHERAL  Likelihood = 3.55    117
modified ALOM score: 1.11

*** Reasoning Step: 3
```

The protein has homology with the following sequences in the databases:

951            981            1011            1041            1071            1098            1128            1158  
HYVEDGNDLMAVYEKMEHVINYVRSGNGPAIVEVESYRWFGHSTADAGVVRTKEE-VDSWKAKDPVKRYRAYLIENEIAT  
|| :|::|||: | | |:| ||::| :|| :|| || :||:| | : : || :: :| :|:: |  
-VTVDGKDILAVYQAAEEAIERARNGGGPSLTIECTMYRNYGHFEGDAQTYKTKDERVEHLLEEKDAIQGFKNYLLKETDAN

210            220            230            240            250            260            270            280

```

1188      1218      1248      1278      1308      1338      1368      1398
EEELAAIEAQVIKEVEEGVKFAEESFPFDMSSVAFEDVFD*NNLK*MRPISFYYSID*KTDIRRK**AKLKLWLCAKRLM
:|:| |:|:|:| |:|:|:| |:|:|:| |:|:|:|
--KLSDIEQVRVSESIKAVSFSEDSPPYKDSSELLTDVYVSYEKGGM
      300      310      320      330

```

GBS403-GST was purified as shown in Figure 218, lane 6.

### Example 1394

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.2464(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12414 GB:Z99107 similar to ABC transporter (ATP-binding
      protein) [Bacillus subtilis]
Identities = 328/643 (51%), Positives = 443/643 (68%), Gaps = 9/643 (1%)
```

Query: 9 MIILQGNKIERFSGDVLFDNINIQVDQRDRIALVGRNGAGKSTLLKILVGEEAPTKEI 68  
M+ILQ N++ +SF D + +NI ++V RDRIA+VGRNGAGKSTLLKI+ G+ + KGEI  
Subject: 1 MMILANOLSKSFGADTILNNIKLEVRNRDRIAIVGRNGAGKSTLLKIIAGQLSYEKGEI 60

Query: 69 NKKRDLSLSYLAQDSRFQSENTTFQEMLQVFDLSREVEKRLRELELQMGQVSGSDLEQLM 128  
K +D+++ YLAQ + S+ TI +E+L VFD L+ +EK +R +E +M +LE +M  
Subject: 61 IKPKDITMGYLAOHTGLDSKLTIKEELLTVFDHLKAMEKEMRAMEEKMAAADPGELESIM 120

Query: 129 KTYDILSEFREKGGFTYESDIKAILNGFKFNSDMWEMPISELGGQNTRLALAKMLLEK 188  
 KTYD L +EF++KGG+ YE+D++++L+G F+ + LSGGQ TRLAL K+LL +  
 Subject: 121 KTYDRLQEEFKDKGGYQYEAADVRSVLHGLGFSHFDDSTOVOSLSGGOKTRLALGKLLLTQ 180

Query: 189 PELLVLDEPTNHLHDIDTTIAWLENLYLVNYQGALIIIVSHDRYFLDKVATVTYDLTHSLDRY 248  
P+LL+LDEPTNHLHDIDT+ WLE+YL Y GA++IVSHDRYFLDKV Y+++ +Y  
Subject: 181 PDLLILDEPTNHLHDIDTLTWLEHYLOGYSGAILIVSHDRYFLDKVNVQVVEVSRAESKKY 240

Query: 249 VGNYSKFMDLKAEKIATEEKNFEKQOKEIAKLEDFVQRNIVRASTTKRAQARRKQLEKME 308  
GNYS ++D KA + + K +EKQQ EIAKL+DFV RN+ RASTTKRAQ+RRKQLE+M+  
Subject: 241 HGNSAYSLDQKAAQYEKDLKMYEKQODEIAKLQDFVDRNLARASTTKRAQSRRKQLERMD 300

Query: 309 RLDKPNVEQKSANMTFHAGKVSNGNVLTLENAAGYEG-VSLSEPIDLDVKKFDALAIVG 367  
+ KP ++KSAN F K SGN VL +++ I YE L + + + ++ A+VG  
Sbjct: 301 VMSKPLGDEKSANFHFIDITKQSGNEVLRVQDLTISYENOPPLTEVSFMLTRGESAAALVG 360

Query: 368 PNGIGKSTLIKSLVGQIPFIKGEAKLGANVETGYDQSQSNLTKTNTVLDELWDAFSTTP 427  
 PNGIGKSTL+K+L+ + +G G+NV GGYDQ O+ LT + VLDELWD + P

-1533-

Sbjct: 361 PNGIGKSTLLKTLIDTLKPDQGTISYGSNVSVGYDQEQAELTSSKRVLDLWDEYPLP 420

Query: 428 EVEIRNRLGAFLFSGDDVKKSVMLSGGERARLLAKLSMENNFLILDEPTNHLDDISK 487  
E EIR LG FLFSGDDV K V LSGGE+ARL LAKL ++ NFLILDEPTNHLDDISK

5 Sbjct: 421 EKEIRTCIGNFLFSGDDVLKPVHSLSGGEKARLALAKMLQKANFLILDEPTNHLDDISK 480

Query: 488 EVLENALIEFDGTLFLVSHDRYFINRVATKVLEISDKGSTLYLGDYDYLLTKKAELEELA 547  
EVLENALI++ GTLLFVSHDRYFINR+AT+VLE+S YLGDYDYY KK E EL

10 Sbjct: 481 EVLENALIDYPGTLLFVSHDRYFINRIATRVLELSSSHIEYLG DYDYTEKKTEQLELE 540

Query: 548 RLNEEEVSASKTEIDVTS-----YETQKANQKEFRKITRRVVEIEARLEVLNENNING 603  
++N++E KT V SD YE +K +K+ R+ RR+ EIE ++ +E + + +

Sbjct: 541 KMNQOE-ETDKTPATVKS DSKRSYEEKEWKKEQRRLRIEETETTQVTEENISRND 599

15 Query: 604 LMLET---NDIGKLSDLQKELESIQEEQLLLMEEWENLNMRLD 643  
L+ + D K+ + + E + +E L+ EWE L+ D

Sbjct: 600 LLCDEPEVYQDHEKVQAIHADNEKLNQEELESLLSEWEELSTEED 642

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4279> which encodes the amino acid  
20 sequence <SEQ ID 4280>. Analysis of this protein sequence reveals the following:

Possible site: 56  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2042(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 473/635 (74%), Positives = 545/635 (85%), Gaps = 1/635 (0%)

Query: 9 MIILQGNKIERFSFGDVLFDNINIQVDQRDRIALVGRNGAGKSTLLKILVGEEAPTKEI 68  
MIILQGNK+ERSFSFGDVL NI++QVD+RDRIALVG NGAGKSTLLK+LVGEE PT GE+

35 Sbjct: 1 MIILQGNKLSRFSFGDVLQFNISLQVDERDRIALVGPNGAGKSTLLKLVGEETPTSGEV 60

Query: 69 NKKRDLSSYLAQDSRFQSENTIFQEMLVFDSLREVEKRLRELELQMGQVSGSDLEQLM 128  
N K+DL+LSYLAQ+SRF+S+ TI++EML+VF++LR+ EKRLR++E+ M VSG L +LM

Sbjct: 61 NTKKDLTSLYLAQNSRFESDQTIYEEMLVFEALRQDEKRLRQMEMDMATVSGQVLTRLM 120

40 Query: 129 KTYDILSEEFREKGGFTYESDIKAILNGFKFNSDMWEMPISLSEGGQNTLALAKMLLEK 188  
YD+L+E FR++GGFTYESDIKAILNGFKF+ MW+M I+ELSGGQNTLALAKMLLEK

Sbjct: 121 TDYDILLTEHFRQGGFTYESDIKAILNGFKFDESMWQMTIAELSGGQNTLALAKMLLEK 180

45 Query: 189 PELLVLDEPTNHLDDITIAWLENYLVNYQGALIIVSHDRYFLDKVATVTYDLTTHSLDRY 248  
PELLVLDEPTNHLDDI+TIAWLENYL NYQGALIIVSHDRYFLDKVATVT DLT + LDRY

Sbjct: 181 PELLVLDEPTNHLDDIETIAWLENYLANYGALIIVSHDRYFLDKVATVTLDTLTPNGLDRY 240

Query: 249 VGNYSKFMDLKAEKIATEEKNFEKQKKEIAKLEDFVQKNIVRASTTKRAQARRKQLEKME 308  
GNYS+FM LKAEK+ EEK F+KQKKEIAKLEDFVQ+KNIVRASTTKRAQARRKQLEK+E

50 Sbjct: 241 SGNYSRFMALKAEKLVAEEKQFDKQKKEIAKLEDFVQKNIVRASTTKRAQARRKQLEKIE 300

Query: 309 RLDKPNVEQKSANMTFHAGKVSNNVLTLENAAGYEGVSLSEPIDLDVKKFDAIAIVGP 368  
RLDKP +KSA+MTFHA K SGNVVL +E AAIGY LSEPI++D+ K DAIA+VGP

55 Sbjct: 301 RLDKPTGGRKSAHMTFHAKEKPSGNVVLVEEAAGYGDQVLEPINVDINKLDAIAVGP 360

Query: 369 NGIGKSTLIKSLVGQIPFIKGEAKLGANVETGYDQSQSNLTKTNTVLDELWDAPSTTPE 428  
NGIGKSTLIK++GQ+P +KG+ K GANVETGYDQ+QS+LT +NTVL+ELW FSTTPE

Sbjct: 361 NGIGKSTLIKSIIGQLPLKQQLKYGANVETGYDQTSHTLSSNTVLEELWQDFSTTPE 420

60 Query: 429 VEIRNRLGAFLFSGDDVKKSVMLSGGERARLLAKLSMENNFLILDEPTNHLDDIDSKE 488  
V+IRNRLGAFLFSGDDVKKSVM+LSGGE+ARLLAKLSMENNFL+LDEPTNHLDDIDSKE

Sbjct: 421 VDIRNRLGAFLFSGDDVKKSVM+LSGGEKARLLAKLSMENNFLVLDEPTNHLDDIDSKE 480

65 Query: 489 VLENALIEFDGTLFLVSHDRYFINRVATKVLEISDKGSTLYLGDYDYLLTKKAELEELAR 548  
VLENALI+FDGTLFLVSHDRYFINR+ATKVLEI++ GSTLYLGDYDYLL KKAEELEELAR

-1534-

Sbjct: 481 VLENALIDFDGTLFVSHDRYFINRLATKVLETTENGSTLYLGDYDYLYLEKKAELEELAR 540

Query: 549 LNEEEVSASKTEIDVTSDYETQKANQKEFRKITRRVVEIEARLEVLNENNINGLMLET 608  
 L E E T DY+ QKANQKE R++TRR EIEARLE +E I M +

5 Sbjct: 541 LAAGETVEETKEASAT-DYQLQKANQKERRRLTRYEEIEARLETIEERIGAIQEDMHAS 599

Query: 609 NDIGKLSDLQKELESTIQEEQLLMEEWENLNMRLD 643  
 ND +L QKE + + +EQ LMEEWE + +++

10 Sbjct: 600 NDTAQLIAWQKEWDQLDQEQEALMEEWETIAEQIE 634

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1395

A DNA sequence (GBSx1480) was identified in *S.agalactiae* <SEQ ID 4281> which encodes the amino acid sequence <SEQ ID 4282>. This protein is predicted to be thiophene degradation protein F (thdF). Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0876(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9773> which encodes amino acid sequence <SEQ ID 9774> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4283> which encodes the amino acid sequence <SEQ ID 4284>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0795(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 384/458 (83%), Positives = 427/458 (92%)

Query: 12 MSITKEFDTIAAISTPLGEGAIGIVRISGTDALKIASKIYRGKDLSAIQSHTLNIGHIVD 71  
 MSITKEFDTI AISTPLGEGAIGIVR+SGTDAL IA +++GK+L + SHT+NYGHI++

Sbjct: 1 MSITKEFDTITAISTPLGEGAIGIVRLSGTDALAIQSVFKGNLEQVASHTINYGHIIN 60

Query: 72 PDKNEILDEVMLGVMLAPKTFTRDVEIEINTHGGIAVTNEILQLILRHGARMAPGEFTK 131  
 P I+DEV+ VMLAPKTFTR+V+EINTHGGIAVTNEILQL++R GARMAPGEFTK

45 Sbjct: 61 PKTGTIIDEVMVSVMLAPKTFTRDVEIEINTHGGIAVTNEILQLILRHGARMAPGEFTK 120

Query: 132 RAFLNGRVDLTQAEAVMDLIRAKTDKAMDIYVQKLDGSLKTLINNTROEILNTLAQVEVN 191  
 RAFLNGRVDLTQAEAVMD+IRAKTDKAM IYVQKLDGSL LIN+TROEILNTLAQVEVN

50 Sbjct: 121 RAFLNGRVDLTQAEAVMDIIRAKTDKAMTIYVQKLDGSLQLINDTROEILNTLAQVEVN 180

Query: 192 IDYPEYDDVEEMTTTLMREKTQEFQALMENLLRTARRGKILREGLSTAIIGRPNVGKSSL 251  
 IDYPEYDDVEEMTT L+REKTQEFQ+L+E+LLRTA+RGKILREGLSTAIIGRPNVGKSSL

Sbjct: 181 IDYPEYDDVEEMTTALLREKTQEFQSLLESLLRTAKRGKILREGLSTAIIGRPNVGKSSL 240

55 Query: 252 LNNLLREEKAIVTDIEGTTTRDVIEEYVNIKGVPLKLVDTAGIRDTDDIVEKIGVERSKKA 311  
 LNNLLRE+KAIVTDI GTTRDVIEEYVNIKGVPLKLVDTAGIR+TDD+VE+IGVERSKKA

Sbjct: 241 LNNLLREDKAIVTDIAGTTTRDVIEEYVNIKGVPLKLVDTAGIRETDDLVEQIGVERSKKA 300

-1535-

Query: 312 LEEADLVLLVLNSSEPIITLQDRSLLELSKESNRIVLLNKTDLPQKIEVNELPKNVIPIISV 371  
 L+EADLVLLVLN+SE LT QDR+LL LS++SNRI+LLNKTDL QKIE+ +LP + IPISV  
 Sbjct: 301 LQEADLVLLVLNASEKLTQDRALLNLSQDSNRILLNKTDLEQKIELEQLPDDYIPISV 360

Query: 372 LENENIDKIEERINDIFFDNAGMVEHDATYLSNARHISLIEKAVDSLKAVNEGLELGMFV 431  
 L N+NI+ IE+RIN +FFDNAG+VE DATYLSNARHISLIEKAV SL+AVN+GL LGMPV  
 Sbjct: 361 LTNQININLIEDRINQLFFDNAGLVEQDATYLSNARHISLIEKAVQSLEAVNDGLALGMFV 420

Query: 432 DLLQVDMTRTWEILGEITGDAAPDELITQLFSQFCLGK 469  
 DLLQVD+TRTWEILGEITGDAAPDELITQLFSQFCLGK  
 Sbjct: 421 DLLQVDLTRTWEILGEITGDAAPDELITQLFSQFCLGK 458

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1396

A DNA sequence (GBSx1481) was identified in *S.agalactiae* <SEQ ID 4285> which encodes the amino acid sequence <SEQ ID 4286>. Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -9.18 Transmembrane 280 - 296 ( 276 - 299)  
 INTEGRAL Likelihood = -4.83 Transmembrane 249 - 265 ( 243 - 266)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4673(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]  
 Identities = 88/306 (28%), Positives = 149/306 (47%), Gaps = 17/306 (5%)

Query: 1 MIVEQKFGNGFTWIN---IEAEQLRTETSEIQAKY-LDSEIITYALDDYERAFMECSHIK 56  
 MI +K NG WI I AE+ T ++ +Y +D +II Y D+ E I  
 Sbjct: 1 MIKPEKTINGTKWIETIQINAEERAT----LEDQYGIDEDIIEYVTDNDESTNYVVD-IN 55

Query: 57 GKEVLTIIFNTIDLKQKESYYETVPMTFCLSHDLITVTRSRNSYMLELLQKYLDNRNPDV 116  
 + L I L + Y T P L L T +S + L LD NP+V  
 Sbjct: 56 EDDQLFIFLAPYALDKDALRYITQPFGLMLHKGVLFTFNQSGIPEVNTALYSALD-NPEV 114

Query: 117 -SPKFLFAALTITKQYFNVVSKIDREKDILNRQLREQT TNKRL LAMS DLETG SVYLLT 175  
 S F+ L + + + I +++++ L++ L +T N L+++S L+ +L +  
 Sbjct: 115 KSVDAFILETLFTVVVSFIPISRAITKKRNYLDKMLNRKTKNSDLVSLSYLQQTTLTFLSS 174

Query: 176 AANQNALVLEQLDVHPSQRFNSEVEKEQLS---DALIEAHQLVSM TQLNSQVLSQLSSTF 232  
 A N L +LD P F +++++ D IE Q+ M ++ +QV+ ++ T  
 Sbjct: 175 AVQTN---LSELDRLPKTHFGVGADQDKIDLFEDVQIEGEQVQRMFEIETQVVDRIDHTL 231

Query: 233 NNVLNNNLNENLTGLNIISINLAIIAAITGFFGMNIP LPLTESRSSWLIVIATSVLLWVI 292  
 N++ NNNLN+ + L I S+ +A+ I+GF+GMN+ LPL + +W++ + SV+L V  
 Sbjct: 232 NSLANNLNNDTMKFLTWSLTMAVPTIISGFYGMNVKLPLAGMQYAWMLTLGISVVLIVA 291

Query: 293 IAQILK 298  
 + +LK  
 Sbjct: 292 MLIMLK 297

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1536-

**Example 1397**

A DNA sequence (GBSx1482) was identified in *S.agalactiae* <SEQ ID 4287> which encodes the amino acid sequence <SEQ ID 4288>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 38
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.1437(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1398**

A DNA sequence (GBSx1483) was identified in *S.agalactiae* <SEQ ID 4289> which encodes the amino acid sequence <SEQ ID 4290>. This protein is predicted to be exonuclease RexA. Analysis of this protein sequence reveals the following:

```

20  Possible site: 52
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.3165(Affirmative) < succ>
25  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9771> which encodes amino acid sequence <SEQ ID 9772> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]
Identities = 522/1211 (43%), Positives = 747/1211 (61%), Gaps = 73/1211 (6%)

35  Query: 28   KRTPEQIEAIYTFGNNVLVSASAGSGKTFVMVERILDKLLRGVPIDSLFISTFTVKAAGE 87
      Sbjct: 5   K TPEQ EAI++ G N+LVSASAGSGKTFVM +RI++K+ +G+ ID LFISTFT KAA E
      KLTPEQNEAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQGIIDRLFISTFTTKAASE 64

      Query: 88   LKERLEKKINESLKSASDDLKQFLTQQLVGIQTADIGTMDAFTQKIVNQYGYTLGISPI 147
      Sbjct: 65   LRMRLERDLKKARQESSDDEEAHRLTLALQNLNADIGTMDSFTQKLTKANFNVRNIDEN 124

      Query: 148  FRILQDKNEQDVIKNEVYADLFSDYMTGKNAAS-----FIKLVKNFSGNRKDSKAFFREMV 202
      Sbjct: 125  FRIL D+ E D+I+ EV+ L Y++ + + F KL+KNFS +R + F+++V
      FRILADQTESDLIRQEVFEQLVESYLSADESLNISKDKFEKLIKNFSDKR-NILGFQKV 183

45  Query: 203  YKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAI PDQEVSSLLNMVMT--TANQLRDLTD 260
      Sbjct: 184  Y +Y F+ +T+NP W++ FLKG +TY +++ D +NV + T +L +
      YTIYRFASATENPISWLENQFLKGFETY---KSLTDLSEDFTVNVKENLLTFFELLEAIS 240

50  Query: 261  QEDYKQLTAKGVPTANYKKHLKIENL-VHWSQDFNLLYGKKGLTNLARDITNVIPSGND 319
      Sbjct: 241  KKDFVTCTAL-----FLSIDTDIRVGSSKDEALSALKKDFSAQKQDL----- 282

      Query: 320  VTVAGVKYPIFKQLHNRIVGLKHLEVIFKYQGESLFLLELLQSFVLD FSEQYLQEKIQEN 379

```

-1537-

V P +L + +KH ++I KYQ ++ + LQ F++DF + YL+ K EN  
 Sbjct: 283 --VGSKSKP--GELRKFFVDKIKHGQLEKYQNQAFEIASDLQKFIDFYKTYLERKKNNEN 338  
 5 Query: 380 AFEFSDIAHFAIQILEENHDIRQLYQDKYHEVMVDEYQDNNHTQERMLELLSNGHNRFMV 439  
 AFE+SDIAHFAI+ILEEN DIR+ ++ Y E+M+DEYQD +HTQERMLELLSNGHN FMV  
 Sbjct: 339 AFEYSDIAHFAIEILEENPDIRENLREHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMV 398  
 10 Query: 440 GDIKQSIYFRQADPQIFNDKYKAYQDNPSQGKLIILKENFRSQSEVLDSTNSVFTHLMD 499  
 GDIKQSIY FR ADP +F +KYK+Y + +LI LKENFRS+ EVL+ TN +F HLMD  
 Sbjct: 399 GDIKQSIYGFRLADPGLFLEKYKSYDQAEENPNQLIRLKENFRSRGEVLNFTNDIFKHLMD 458  
 15 Query: 500 EEVGDILYDESHQKAGS---PROQERHPNNKTQVLLLDTDEDDIDDSQDYDISPAE 555  
 E++G++ Y + L G+ P + E+ + + +T E++I+DS+ + IS E  
 Sbjct: 459 EKLGEPTYGKEEALVQGNISDYPVEAEKDFYPELILLYKENTSEEEIEDSEVK---ISDGE 515  
 20 Query: 556 AKLVAKEIIRLHKEENVFPQDITLLVSSRTRNDGILQTFDRYGIPLVTDGGEQNYLKSVE 615  
 K A+EI +L E V +DI +LV S++ N+ I Y IP+V D G ++LKS+E  
 Sbjct: 516 IKGAAQEIKKL-IEYGVEPKDIALIVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSME 574  
 25 Query: 616 VMVMLDTLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQDVK-MAFYHKVKLSYHKE 674  
 V++MLD LR+IDNPL D +LVA+LRSP+FGFNED+LTRI++Q + + F+ K+ Ls KE  
 Sbjct: 575 VLIMLDVLRADNPLYDLSLVAMLRSP+FGFNEDLTRISVQGSRLRFRWDKILLSLKKE 634  
 30 Query: 675 GHHSDLITPELSSKIDHFMKTFQTWRDFAKWHSLYDLIWKIYNDRFYDYVGALPKAEQR 734  
 G + +LI L K+ F + F WR ++ L+WKIY + +Y+DYVGAL E R  
 Sbjct: 635 GKNPELINLSLEQKLKAFNQKFTWRKLVNKIPIHRLWKIYTETYYFDYVGALKNGEMR 694  
 35 Query: 735 QANLYALALRANQFEKTGFKLSRFIRMIDKVLNENDLADVEALPQNAVNLMTIHKSK 794  
 QANL AL++RA +E +G+KGL +F+R+I+K +E NDLA V + LPQNAV +MT HKSK  
 Sbjct: 695 QANLQALSVAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLQNAVVRMTFHKSK 754  
 40 Query: 795 GLEFKYVFILNIDKKFSMVDITSPLILSRNQGIGIKYVADMRHELEE-EILPAVKVSMET 853  
 GLEF YVF++N+ +F+ D+ +ILSR G+G+KY+AD++ E + P V MET  
 Sbjct: 755 GLEFDYVFLMNLQSRFNDRDLKEDVILSREHGLGMKYIADLKAEPDVTITDFPYALVKMET 814  
 45 Query: 854 LPYQLNKRELRLATLSEQMRLLYVAMTRAEEKLYLVGKASQT---KWADHYDLVS-ENNH 909  
 PY +NK + A LSE+MR+LYVA TRA+KKLYLVGK T + YD + E  
 Sbjct: 815 FPYMVNKDLQRAALSEEMRVLYVAFTRAKKLYLVGKIKDTDKKAGLELYDAATLEGKI 874  
 50 Query: 910 LPLASRETFVTFQDWLLAVHETYKKQELFYDINFVSLEELTDHHIGMVNPSLPFNPDK- 968  
 L R + FQ W+LA+ K L +N + +EL + + PD K  
 Sbjct: 875 LSDKFRNSSRGFQHWILALQNAK---LPMKLNVTYKDELETEKLEFTS-----QPDFKK 926  
 55 Query: 969 -VENRQSEDIVRAIS--VLESVEQINQTY--KAAIELPTVTRTPSQVKK-IYEPILDIEGV 1022  
 VE + D + + S + E+ + +N Y +AA EL +++TPSQVKK YE L + V  
 Sbjct: 927 LVEESEKFDNIMSFSEIKEAQKIMNYQYPHQAATELSSIQTPSQVKKRSYEKQLQVGEV 986  
 60 Query: 1023 D-VMETITKTSVDFKLDPDFSTSKKQDPAALGSVHELMQRIEMSSHVKMEDIQKALTEVN 1081  
 V E + ++DF DF KK A +GSA H MQ + S + Q L E+  
 Sbjct: 987 QPVSEFVRVKNLDFS--DFG-PKKITAAEMGSATHSFMQYADF-SQADLFSFQATLDEM 1042  
 65 Query: 1082 AETSVKAAIQIEKINYFFQETSLGKYIQEEVHLHREAPFAMLKEDPESGEKFFVVRGIID 1141  
 + +K I I KI F +T G+++ E V+ +EAPF+ML+ D + E+++VRGI D  
 Sbjct: 1043 FDEKIKNQIDITKILTLF-DTEFGQFLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICD 1101  
 70 Query: 1142 GYLLLENRIILFDYKTDKFNVP---LELKERYQGQMALYAEALKKSYETIEKIDKYLILLG 1198  
 G++ L ++IILFDYKTD+F N E+KERY+ QM LY+EAL+K+Y + +IDKYLILLG  
 Sbjct: 1102 GFVKLADKIILFDYKTDRTNVSASEIKERYKDQMNLYSEALQAYHVNQIDKYLILLG 1161  
 75 Query: 1199 G-KQLEVVKMD 1208  
 G +++ V K+D  
 Sbjct: 1162 GPRKVFVEKID 1172

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4291> which encodes the amino acid  
 sequence <SEQ ID 4292>. Analysis of this protein sequence reveals the following:

Possible site: 61

-1538-

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

----- Final Results -----

5                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

10 >GP:AAC12966 GB:U76424 exonuclease RexA [Lactococcus lactis]  
 Identities = 478/1206 (39%), Positives = 700/1206 (57%), Gaps = 65/1206 (5%)

Query: 40   KRTAQQIEAIYTSQGNILVSASAGSGKTFVMVERILDKILRGVSIDRLFISTFTVKAATE 99  
           K T +Q EAI++SG+NILVSASAGSGKTFVM +RI++K+ +G+ IDRLFISTFT KAA+E  
 Sbjct: 5   KLTPQEAEIAIHSSGKNILVSASAGSGKTFVMAQRIVEKVKQGIEIDRLFISTFTKKAASE 64

15 Query: 100   LRERIENKLYSQIAQTTFQFMKVLTEQLQSLCQADIGTMDAFAQKVVSRYGYSIGISSQ 159  
           LR R+E L   +++D +   LT LQ+L ADIGTMD+F QK+   + I  
 Sbjct: 65   LRMRLELDLKKARQESSDDEEAHRLTLALQNLNADIGTMDSFTQKLTKANFNVRNIDPN 124

20 Query: 160   FRIMQDKAEQDVLKQEVFSKLFNEFMNQKEA-----PVFRALVKNFSGNCKDTSAFRELV 214  
           FRI+ D+ E D+++QEVF +L   +++ E+   F L+KNFS + ++ F+++V  
 Sbjct: 125   FRILADQTESDLIRQEVFEQLVESYLSADESLNISKDKFEKLIKNFSDK-RNILGFQKVV 183

25 Query: 215   YTCYSFSQSTENPKIWLQENFLSAAKTYQRLEDIPDHDIELLLAMQDTANQLRDVTDME 274  
           YT Y F+ +TENP WL+ FL   +TY+ L D+ + D + +   T +L +   +  
 Sbjct: 184   YTIYRFASATENPISWLENQFLKGFETYKSLTDLSE-DFTVNVKENLLTFFELLEAISK 242

30 Query: 275   DYGQLTKAG-SRSAKYTKHLTIIEKLSDWVRDFKCLYGKAGLDRLIRDVTGLIPSGNDVT 333  
           D+ T S                   + E LS   +DF                   D+  
 Sbjct: 243   DFVCTALFLSIDTDIRVGSSKDEALSALKKDFSA-----QKQDLV 283

35 Query: 334   VSKVKYPVFKTLHQKQFRHLETILMYQKDCFSLLQLQDFVLAFSEAYLAVKIQESAF 393  
           SK K   +   K+K H + I YQ F + LQ F++ F + YL K E+AF  
 Sbjct: 284   GSKSKPGELRKFDKIK---HGQLIEKYQNQAFELASDLQKFIIDFYKTYLERKKENAF 340

40 Query: 394   EFSIDIAHFAIKILEENTDIRQSYQQHYHEVMVDEYQDNNHMQERLLTLLSNGHNRFMVGD 453  
           E+SDIAHFAI+ILEEN DIR++ ++HY E+M+DEYQD +H QER+L LLSNGHN FMVGD  
 Sbjct: 341   EYSDIAHFAIEILEENPDIRENLRREHYDEIMIDEYQDTSHTQERMLELLSNGHNLFMVGD 400

45 Query: 454   IKQSIYFRQADPQIFNQKFRDYQKKPEQKGVILLKENFRSQSEVLNVSNVAFSHLMDES 513  
           IKQSIY FR ADP +F +K++ Y +   ++I LKENFRS+ EVLN +N +F HLMDE  
 Sbjct: 401   IKQSIYGFRLADPGLFLEKYKSYDQAEENPNQLIRLKENFRSRGEVLNFTNDIFKHLMDEK 460

50 Query: 514   VGDVLYDEQHQLIAG--SHAQTPPYLDRAQLLLYNSDKDDGNAPSDSEGISFSEVTIVA 571  
           +G++ Y ++ L+ G S                   D +LLLY + +                   IS E+ A  
 Sbjct: 461   LGEMTYGKEEALVQGNISDYPVEAEKDFYPELLLYKENTSEEBIEDSEVKISDGEIKGAA 520

55 Query: 572   KEIIKLHNDKGVPFEDITLLVSSRTRNDIISHTFNQYGIPIATDGGQQNYLKSVEVMVML 631  
           +EI KL + GV +DI +LV S++ N+ I                   Y IP+ D G+ ++LKS+EV++ML  
 Sbjct: 521   QEIKKL-IEYGVPEPKDIALIVRSKSNNNKIEDILLSYDIPVVLDEGRVDFLKSMEVLIML 579

60 Query: 632   DTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDNELDKDCLYDKIQRAVIGRGA 691  
           D LR I+NP D +LVA+LRSP+F F+ED+L RI++Q   +L   +DKI ++ G  
 Sbjct: 580   DVLRAIDNPLYDLSLVAMLRSPFGFNEDELTRISVQGSRDLR---FWDKILLSLKKEGK 636

65 Query: 692   HPELIHDTLLGKLVFLKTKSWRRYAKLGSYDLIWKIFNDRFYDFDVASQAKAEQAQA 751  
           +PELI+ +L KL F +   WR+   ++ L+WKI+ + +YFD+V +   E QA  
 Sbjct: 637   NPELINLSLEQKLKAFNQKFTWRKLVNKIPIHRLLLWKIYTYFYFDYVGALKNGEMRQA 696

Query: 752   NLIALALRANQFEKSGYKGLYRFIKMIDKVLQNDLADVEVATPKQAVNLMTHKSKGL 811  
           NL AL++RA +E SGYKGL++F+++I+K +E NDIA V + P+ AV +MT HSKSKGL  
 Sbjct: 697   NLQALSVRAESYESSGYKGLFKFVRLINKFMEQNNDLASVNIKLPQNAVVRVMTFHKSKGL 756

Query: 812   QFPYVFILNCDKRFSMTDIHKSFILNRQHGIGIKYLADIKGLLGE-TTLNSVKVSMETLP 870  
           +F YVF++N RF+ D+ + IL+R+HG+G+KY+AD+K   T   V MET P  
 Sbjct: 757   EFDYVFLMNLQSRFNDRLKEDVILSREHGLGMKYIADLKAEPDVTDFPYALVKMETFP 816

-1539-

Query: 871 YQLNKQELRLATLSEEMRLLYVAMTRAEEKVYFIGK--ASKSKSQEITDPKKL-GKLLP 926  
 Y +NK + A LSEEMR+LYVA TRA+KK+Y +GK K E+ D L GK+L  
 Sbjct: 817 YMVNKKDLKQRAALSEEMRVLYVAFTRAKKLYLVGKIKDTPDKKAGLELYDAATLEGGKILS 876

5 Query: 927 LALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGRQLQTP---QLLNPPDDL 983  
 R FQ W+LA+ + L + +L E + P +L+ +  
 Sbjct: 877 DKFRNSSRGRFQHWILALQ---NATKLPKMLNVYTKDELETEKLEFTSQPDFKKLVEESEK 933

10 Query: 984 KDNQSEETIARALDMLAVALSOLNANY--EAAIHLPTVRTPSQL-KATYEPLLEPIGVDI 1040  
 DN S + ++ EA +N Y +AA L +++TPSQ+ K +YE L+ V +  
 Sbjct: 934 FDNIMSFSD---EIKEAQKIMNYQYPHQAATELSSIQTPSQVKKRSYEQQLQVGEVQPV 989

15 Query: 1041 EKSSRSLSDFTLPHFSKKAKVEASHIGSALHQLMQVPLSKP--INQOTLLDALRGIDSN 1098  
 + R + + F K K+ A+ +GSA H MQ S+ + Q LD + G D  
 Sbjct: 990 SEFVR-VKNLDFSDFGPK-KITAAEMGSATHSFMQYADFSQADLFSFQATLDEM-GFD-- 1044

20 Query: 1099 EEVKTALDLKKIESFFCDTSLGQFFQTYQKHYREAPFAIKLDPISQEEYVLRGIIDAY 1158  
 E++K +D+ KI + F DT GQF +EAPF++L+ D ++E+Y++RGI D +  
 Sbjct: 1045 EKIKNQIDITKILTLF-DTEFGQFLSENVDKTVKEAPFSMLRTDEFAKEQYIVRGICDGF 1103

25 Query: 1159 FLFDDHIVLVYDKTDKYQP---IELKKRYQQOLELYAEALTQTYKLPVTKRYLVLMGGG 1215  
 D I+L DYKTD++ E+K+RY+ Q+ LY+EAL + Y + +YL+L+GG  
 Sbjct: 1104 VKLADKIILFDYKTDRTNVSASEIKERYKQMNLYSEALQKAYHVNQIDKYLILLGGP 1163

Query: 1216 KPEIVE 1221  
 + VE  
 Sbjct: 1164 RKVFVE 1169

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 728/1211 (60%), Positives = 916/1211 (75%), Gaps = 5/1211 (0%)

Query: 1 MMTFKPFLNPEDIAVIQTEEKNSDKKQKRTPEQIEAIYTFGNVLVSASAGSGKTFVMVE 60  
 +++F PFL+PE I +Q E+ D+ QKRT +QIEAIYT G N+LVSASAGSGKTFVMVE  
 Sbjct: 13 VISFAPFLSPEAIKHLQENERCDSQKRTAQIEAIYTSQGNILVSASAGSGKTFVMVE 72

35 Query: 61 RILDKLLRGVPIDSLFISTFTVKAAGELKERLEKKINESLKSASDDLKQFLTQQLVGIQ 120  
 RILDK+LRGV ID LFISTFTVKA EL+ER+E K+ + +K +LT+QL +  
 Sbjct: 73 RILDKILRGVSDRLFISTFTVKAATELRERIEKNLYSQAQTTFQMKVYLTEQLQSLC 132

40 Query: 121 TADIGTMDAFTQKIVNQYGYTLGISPIFRILQDKNEQDVIKNEVYADLFSDYMTGKNAAS 180  
 ADIGTMDAF QK+V++YGY++GIS FRI+QDK EQDV+K EV++ LF+++M K A  
 Sbjct: 133 QADIGTMDAFAQKVVSRYGYSIGISSQFRIMQDKAEQDVLKQEVFSKLFNEFMNQKEAPV 192

45 Query: 181 FIKLVKNFSGNRKDSKAFREVMYKVYAFSQSTDNPKRWMQTVFLKGAQTYTDFEAI PDQE 240  
 F LVKNFSGN KD+ AFRE+VY Y+FSQST+NPK W+Q FL A+TY E IPD +  
 Sbjct: 193 FRALVKNFSGNCKDTSAFRELVTYCSFSQSTENPKIWLQENFLSAAKTYQRLEDIPDHD 252

50 Query: 241 VSSLLNVMTTANQLRDLTDQEDYKQLTAKGVPTANYKHLKIIENLVHWSQDFNLLYGK 300  
 + LL MQ TANQLRD+TD EDY QLT G +A Y KHL IIE L W +DF LYGK  
 Sbjct: 253 IELLLLAMQDTANQLRDVTDMDYDGLTKAGSRSKATYKHLTIIEKLSDWVRDFKCLYGK 312

55 Query: 301 KGLTNLARDITNVIPSGNDVTVAGVKYPIFKQLHNRIVGLKHLEVIFKYQGESLFLLELL 360  
 GL L RD+T +IPSGNDVTV+ VKYP+FK LH ++ +HLE I YQ + LLE L  
 Sbjct: 313 AGLDRLIRDVTGLIPSGNDVTVSKVKYPVFKTLHQKLKQFRHLETILMYQKDCFSLEQL 372

60 Query: 361 QSFVLDSEQYLQEKIQENAFESDIAHFQIILENHDIRQLYQDKYHEVMVDEYQDNN 420  
 Q FVL FSE YL KIQE+AFESDIAHFQI+ILEEN DIRQ YQ YHEVMVDEYQDNN  
 Sbjct: 373 QDFVLAFSEAYLAVKIQESAFESDIAHFQIILEENTDIRQSYQQHYHEVMVDEYQDNN 432

65 Query: 421 HTQERMLELLSNGHNRFMVGDIKQSIYRFRQADPQIFNDKYKAYQDNPSQGLIILKENF 480  
 H QER+L LLSNGHNRFMVGDIKQSIYRFRQADPQIFN K++ YQ P QGK+I+LKENF  
 Sbjct: 433 HMQERLLTLLSNGHNRFMVGDIKQSIYRFRQADPQIFNQKFRDYQKKPEQGVILLKENF 492

Query: 481 RSQSEVLSTNSVFTHLMDDEEVGDILYDESHQLKAGSPRQQRHPNNKTQVLLLDTDDEDD 540  
 RSQSEVL+ +N+VF+HLMDE VGD+LYDE HQL AGS Q + + + Q+LL ++D+DD  
 Sbjct: 493 RSQSEVLNVSNAVFSHLMDESVDVLYDEHQLIAGSHAQTVPYLDRRAQLLLYNSDKDD 552

-1540-

Query: 541 IDSDSQYVDISPAEAKLVAKIIRLHKEENVFPQDITLLVSSRTRNDGILQTFDRYGIP 600  
 ++ S IS +E +VAKEII+LH ++ VPF+DITLLVSSRTRND I TF++YGIP  
 Sbjct: 553 -GNAPSDSEGISFSEVTIVAKEIIKLHNDKGVFPFEDITLLVSSRTRNDIISHTFNQYGIP 611

5 Query: 601 LVTDGGEQNYLKSVEVMVMDTLRSIDNPLNDYALVALLRSPMFGFNEDDLTRIAIQD-- 658  
 + TDGG+QNYLKSVEVMVMDTLR+I+NP NDYALVALLRSPMF F+EDDL RIA+Q  
 Sbjct: 612 IATDGGQNYLKSVEVMVMDTLRTINNPRNDYALVALLRSPMFAFDEDDLARIALQKDN 671

10 Query: 659 --VKMAFYHKVLSYHKEGHSDLTIPELSSKIDHFMKTFQTWRDFAKWHSLYDLIWKIY 716  
 K Y K++ + G H +LI L K++ F+KT ++WR +AK SLVDLIWKI+  
 Sbjct: 672 ELDKDCLYDKIQRAVIGRGAPHELIHDITLLGKLNVLKTLKSWRRYAKLGSYDLIWKIF 731

15 Query: 717 NDRFYDYVVGALPKAEQRQANLYALALRANQFEKTGFKGLSRFIRMIDKVLENENDLADV 776  
 NDRFY+D+V + KAEQ QANLYALALRANQFEK+G+KGL RFI+MIDKVLE +NDLADV  
 Sbjct: 732 NDRFYDFDVASQAKAEQAQANLYALALRANQFEKSGYKGLYRFIKMIDKVLETQNDLADV 791

20 Query: 777 EVALPQNAVNLMITHKSKGLEFKYVFILNIDKKFSMDITSPILSRNQGIGIKYVADMR 836  
 EVA P+ AVNLMITHKSKGL+F YVFILN DK+FSM DI IL+R GIGIKY+AD++  
 Sbjct: 792 EVATPKQAVNLMITHKSKGLQFPYVFILNCDKRFSMTDIHKSFILNRQHGIGIKYLADIK 851

25 Query: 837 HELEEEILPAVKVSMETLPYQLNKRELRLATLSEQMRLLYVAMTRAEEKLYLVGKASQTK 896  
 L E L +VKVSMETLPYQLNK+ELRLATLSE+MRLLYVAMTRAEEK+Y +GKAS++K  
 Sbjct: 852 GLLGETTLNSVKVSMETLPYQLNKQELRLATLSEEMRLLYVAMTRAEEKVYFIGKASKSK 911

30 Query: 897 WADHYDLVSENNHLEPLASRETFVTFQDWLLAVHETYKKQELFYDINFVSLEELTDHHIGM 956  
 + D LPLA RE +TFQDWLLA+ + + ++L++D+ F+ +LT +G  
 Sbjct: 912 SQEITDPKLGKLLPLALREQLLTFQDWLLAIADIFSTEDLYFDVRFIEDSDLTQESVGR 971

35 Query: 957 VNPSLPFNPNDNKVENRQSEDIVRAISVLESVEQINQTYKAAIELPTVRTPSQVKKIYEPI 1016  
 + NPD+ +NRQSE I RA+ +LE+V Q+N Y+AAI LPTVRTPSQ+K YEP+  
 Sbjct: 972 LQTPQLLNPDLDKDNRQSETIARALDMLAVALNANYEAAIHLPTVRTPSQLKATYEPI 1031

40 Query: 1017 LDIEGVDVMEITITKTSVDFKLPDFSTSKKQDPAALGSVAHELMQRIEMSSHVKMEDIQKA 1076  
 L+ GVD++E +++ DF LP FS K + + +GSA+H+LMQ + +S + + + A  
 Sbjct: 1032 LEPIGVDIIEKSSRSLSDFTLPHFSKAKVEASHIGSALHQLMQVPLPSKPINQQTLLDA 1091

45 Query: 1077 LTEVNAETSVKAAIQIEKINYFFQETSLGKYIQEEVEHLHREAPFAMLKEDPESGEKFVV 1136  
 L +++ VK A+ ++KI FF +TSLG++ Q +HL+REAPFA+LK DP S E++V+  
 Sbjct: 1092 LRIGIDSNEEVKTALDLKKIESFFCDTSLGQFFQTYQKHLYREAPFAIKLDPISQEEYVL 1151

50 Query: 1137 RGIIDGYLLENRIILFDYKTDKFNPLELKERYQGQMALYAEALKKSYETIEKIDKYLIL 1196  
 RGIID Y L ++ I+L DYKTDK+ P+ELK+RYQ Q+ LYAEAL ++Y++ +YL+L  
 Sbjct: 1152 RGIIDAYFLFDDHIVLVDYKTDKYQPIELKKRYQQQLELYAEALTQTYKLPVTKRYLVL 1211

45 Query: 1197 LGGKQLEVVKM 1207  
 +GG + E+V++  
 Sbjct: 1212 MGGGKPEIVEV 1222

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1399

A DNA sequence (GBSx1484) was identified in *S.agalactiae* <SEQ ID 4293> which encodes the amino acid sequence <SEQ ID 4294>. This protein is predicted to be exonuclease RexB. Analysis of this protein sequence reveals the following:

55 Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0660(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1541-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC12965 GB:U76424 exonuclease RexB [Lactococcus lactis]  
Identities = 363/1093 (33%), Positives = 604/1093 (55%), Gaps = 67/1093 (6%)

5 Query: 1 MKLLYTDINHDMTEILVNQAAHAAEAGWRIFYIAPNSLSFEKERAVLENLPQ---EASFA 57  
M++LYT+I D+TE L+ A E +++YI P+S+SFEKE+ +LE L + A F  
Sbjct: 1 MEILYTEITQDLTEGLLEIALEELEKRNKVVYIVPSSMSFEKEKEILERLAKGSDTAVFD 60

10 Query: 58 ITITRFAQLARYFTLNQP-NQKESLNDIGLAMIFYRALASFEDGQLKVFGRLKQDASFIS 116  
+ +TRF QL YF + K L +GL+M+R L SF+ ++ ++ L+ A F+  
Sbjct: 61 LLVTRFQQLPYFDKREKATMKTGTVGLSMLFRRVLRSPFKKDEIPLYFSLQDSAGFLE 120

15 Query: 117 QLVLDYKELQTANLSILELKYLHSPEKFEDLLAIFLVVSDLLREGEYDNQSKIAFFTEQV 176  
L+ L EL TANLS+ L ++ + +LA F + EY N S+ FT ++  
Sbjct: 121 MLIQLRAELLTANLSVENLPDNPKNQELKKILAKFEAELSV----EYANYSEFGDFTNRL 176

20 Query: 177 RSGQLDVLKNTILIVDGFTRFSAEELIKSLSSRCQEIIIGAYASQKAYKANFTNGNI 236  
G+ D LK+ +I+DG+TRFSAEEL I+S+ + ++G Y+ + + A + I  
Sbjct: 177 VDGEFDQQLKDVITIIIDGYTRFSAEELFIESIQEKVARFVVGTYSDENSLTAG--SETI 234

25 Query: 237 YSAGVDFRLYLATTFQTKPEFILSKWESKSGFEMISK-----NIEGKHDFTNSSHILDDT 291  
Y + T F+ K L K S + E+ SK +++ + T+ L  
Sbjct: 235 YVGTSQMI----TRFRNKFPVELRKIASSAVNEVYSKLTRILDLDLSRFVITDEKIELKAE 290

30 Query: 292 AKDCITIWECINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDSYKLQLSKIFEQYDI 351  
+ IWE NQK E+E VA+ IRQK+ QG +KD VL+GD +Y++ L ++F+ Y+I  
Sbjct: 291 DEKYFRIWEAENQKVEIERVAKAIRQKIIQGAFFKDFTVLVGDPAAYEITLKEVFDLYEI 350

35 Query: 352 PYYFGKAETMAAHPLVHFMDLSRIKRYRFRAEDVLNLFKTGIYGEISQDD--LDYFEAY 409  
P+++ + E+M+ HPLV F +SL IK+ +R +DV+NL K+ +Y + + D+ +DYFE Y  
Sbjct: 351 PFFYAQEESSQHPLVIFVFESLFAIKKNYRTDDVNNLLKSKVYTDANLDEEVIDYFEY 410

40 Query: 410 ISYADIKGPKKFFTFVVGAKKFDLGRNLTIRQSLL---TPLESFV-KTKKQDGIKTLNQ 465  
+ I G KKF +F+ ++ + +N +R+ LL +PL+ F+ +K+ G K ++  
Sbjct: 411 VQKYKISGRKKFTEEFIE-SEFSQIELVNEMREKLLGSESPLQVFLGNNRKKTKGKVVSD 469

45 Query: 466 FMFFLTQVGLSDNLSRLVGQMS-ENEQE---KHQEVWKTFTDILEQFQTFGQEKLNLD 521  
L + N++ +NE + KH++VW+ L +F +F EKL E  
Sbjct: 470 LQGLLENGNVMNMYFSAELQNEHQADKHEQVWQMLISTLNEFLAVFSDEKLKSVE 529

50 Query: 522 FLSLLNSGMMQAEYRMVPATVDVVTVKSYDLVEPHSNQFVYALGMTQSHFPKIAQNKSLI 581  
FL +L +G+ A+YR +PA VDVV VK Y+LVEP +N+++YA+G++Q++FP+I +N +L+  
Sbjct: 530 FLDILLAGLKNKAKYRQIPANVDVNVKDYELVEPKTNKYIYAIGLSQTNFPRIKKNSTLL 589

55 Query: 582 SDIERQLINDANDTDGHFDIMTQENLKNHFAALSFLNAKQELVLTIPQLLNESEDQMS 641  
SD ER IN D + + N +KN F LSL N+AK+ LVL++PQ++ + + S  
Sbjct: 590 SDEERLEINQTTDENQFIEQLNVANYQKNQFTVLSLINSAKESLVLSPQIMANEQGEFS 649

60 Query: 642 P-YLVELRDIGVFPNHKGR-QSLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEE-QTF 698  
P + + L+D K + +L E ++IGN +++++ + + R ++ E T E+ + F  
Sbjct: 650 PVFQLFLKDADEKILQKIQGVNLFESLEHIGNSRSVIAMIGQIERELVESEETSEDKRVF 709

65 Query: 699 WSVAVRYLRRQLTSKGIEIPIITDSDLDTVTSSDVMTRFPEDDPLKLSSSALTTFYNNQ 758  
WS R L + + + +D+TV ++ D + + + D + S S+ FYN +  
Sbjct: 710 WSSIFRILVKSNAFQKILDLAKDIDTVNLAPDTLEQIY--GDKIYASVSSFERFYNCE 767

70 Query: 759 YKYFLQYVVLGLEEQDSIHDPMRHHGTYLHRVFEILMKNQGI--ESFEEKLNSAINKTNOE 816  
Y+YFL+ L LE ++I + + G + H VFE +MK + E+F+EKL + + ++  
Sbjct: 768 YQYFLENTLSLETFFENIDINSKI VGNFFHEVFEKVMKETDLAENFDEKLTLLVLQEV DKN 827

75 Query: 817 DVFKSLYSEDAESRYSLEILEDIARATATILR---QDSQMTVESE-----EERFELM 865  
+ +++DA +R++ LE+I R TAT+L+ D T+ +E E  
Sbjct: 828 --YSRYFTQDATARFTWSNLEEIVRQTATVLKATVSTDELKTLTLESSFGLPKSELGNFS 885

80 Query: 866 IDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKFDIQKFYNGLSPLVTYIDAISRDKEV 925  
+D+ I + G IDR+D+LS LG +DYKSSA F +Q+ Y+GLS Q +TY+D I K+  
Sbjct: 886 VDD-IYLRGRIDRLDQLSTDYLGAI DYKSSAHSFKLQEA YDGLSLQFMTYLDVI---KQA 941

-1542-

Query: 926 EQKPPIFGAMYLHMQEPQDLSKIKNLDDLVTKNHQALTYKGLFSEAEKEFLANGKYHL- 984  
 I+GA+YL + +LS+I L ++ +++ Y+GL E E + G ++  
 Sbjct: 942 FPNQKIWGALYLQFKNQPINLSEINQLSEIANILKESMRVEGLVLEDAAEQI-KGIENIA 1000

Query: 985 --KDSLYSETETAILQAHNQSLYKKASETIKSGKFLINPYTEDAKTVGDG-----Q 1033  
 K ++Y+E E L N+ Y+ A + +K GK INP + ++ +D  
 Sbjct: 1001 LKKTNIYNEEEEFEQLLKLNEEHYRAAGQRLKKGKIAINPIMKRSEGIDQSGNVRGCRYCP 1060

Query: 1034 FKSITGFADRHM 1046  
 KSI FEA+ HM  
 Sbjct: 1061 LKSICRFEANIHM 1073

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4295> which encodes the amino acid  
 sequence <SEQ ID 4296>. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1891(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 546/1075 (50%), Positives = 758/1075 (69%), Gaps = 11/1075 (1%)

Query: 1 MKLLYTDINHDMTTEILVNQAAHAAEAGWRIFYIAPNSLSFEKERAVLENLPQEASFAITI 60  
 MKL+YT++++ MTEILVN+A AA+ G+R+FYIAPNSLSFEKER VL LP+ +F+I +  
 Sbjct: 1 MKLIYTEMSSYMTTEILVNEARKAADQGYRVFYIAPNSLSFEKEREVLTLPERGTFSIIV 60

Query: 61 TRFAQLARYFTTLNQPNQKESLNDIGLAMIFYRALASFEDGQLKVFGRLKQDASFISQLVD 120  
 TRF Q++RYFT+ K+ L+D LAMIFYRAL + L +GRL+ ++ FI QLV+  
 Sbjct: 61 TRFVQMSRYFTVESSPSKQHLDDTTLAMIFYRALMQLKPEDLPSYGRLONNSVFIEQLVE 120

Query: 121 LYKELQTANLSILELKYLHSPEKFEDLLAIFLVVSDLLREGEYDNQSKIAFFTEQVRSGQ 180  
 LYKEL+ A LS+ +L L P+K EDL+ I + ++ +Y+ S + F ++ G  
 Sbjct: 121 LYKELKNAQLSVHDLTGLDHPQKQEDLIKIIELAETIMIQQDYNQDSPLQSFARAIKGL 180

Query: 181 LDVDLKNLTILVDGFTFRSAEEALIKSLSSRCQEIIIGAYASQKAYKANFTNGNIYSAG 240  
 L+ L T++++DGF+RFSAAE+ L+ L++ CQE+IIG+Y SQKAY+ +F GNIY A  
 Sbjct: 181 LNNQLSKTVVVIDGFSRFSABEDYLLSLNNNCQEVIIIGSYVSQKAYQKSFIGNIYEAS 240

Query: 241 VDFRLYLATTFQTKPEFILSKWESKSGFEMISKNIEGKHDFTNSSHILDDTAKDCITWE 300  
 + FL+ LA + KP F S K F +++ E HDF+ L + D ++W+  
 Sbjct: 241 LHFLQDLAQKYHIKPVFATSNQVFKPAFSRLTQLFEATHDFSQVDWQLQKSDLDHFSWLQ 300

Query: 301 CINQKDEVEHVARAIRQKLYQGYRYKDILVLLGDVDSYKLQLSKIFEQYDIPYFGKAET 360  
 C +QK+E+EHVA++IRQKLY+GYRYKDILVLLGD+D+Y+LQ+ IF++++IPY GKA  
 Sbjct: 301 CHHQKEEIEHVAKSIRQKLYEGYRYKDILVLLGDMDAYQLQIGPIFDKFEIPYLGKAEP 360

Query: 361 MAAHPLVHFMDLSRIKRYRFRADVLNLFKTGIYGEISQDDLDYFEAYISYADIKGPKK 420  
 MAAHPLV F++SL R +RY +R ED+LN+ K+G++G D+D FE Y +ADIKG K  
 Sbjct: 361 MAAHPLVQFIESLERSQRYNWRREDILNMLKSLFGCFDDSDIDRFEEYTQFADIKGFTK 420

Query: 421 FFTDFVVGAKKFDLGRNLTIROSLTPLESFVTKKQDGIKTLNQFMFFLTQVGLSDNL 479  
 F F + +++ L LN +RQ ++ PL+ K++KQ G +++ + FL ++ L++N+  
 Sbjct: 421 FSKPFTINSSRQYPLDFLNEMRQDIVLPLQELFKSQKQLGASLVDKLLFLKKIRLAENM 480

Query: 480 SRLVGQMSENEQEKHQEVWKTFTDILEQFQTFGQEKLNLDLFLSLLNSGMMQAEYRMVP 539  
 L S+ E EK++EVWK FTDIL F IFGQEK L + L+L+ +GM A+YR+VP  
 Sbjct: 481 QGLA--QSQLEVEKNEEVWKRFTDILTSFHHIFGQEKRLSDCLALIKTGMKSAQYRVVP 538

Query: 540 ATVDVVTVKSYDLVEPHSNQFVYALGMTQSHFPKIAQNKSLISDIERQLINDANDTDGHF 599  
 AT+DVVT+KSYDLV+PHS FVYA+G+TQSHFPK + L+SD ER IN+ + HF  
 Sbjct: 539 ATLDVVTIKSYDLVQPHSKPFVYAIGLTQSHFPKQIHHSGLLSDQERARINEIRNY-RHF 597

-1543-

Query: 600 DIMITQENLKKNHFAALSLFNAAQELVLTIPQLLNESEDQMSPYLVELRDIGVFPNHKGR 659  
 DI + EN KKNH ALSLFNAA +ELVL++ ++NE+ D +SPYL EL + G+P KG+  
 Sbjct: 598 DIASAENSKKNHQTALSLFNAAATKELVLSVSTVINETFDDLSPYLKELINFGPLPLDKGK 657

Query: 660 QSLKEEADNIGNYKALLSRVVDLYRSAIDKEMTKEEQTFWSVAVRYLRRQLTSKGIEIPI 719  
 L + +IGNYKALLS+++ + R + EM+ +++ FW+V +RYLR+QL + +E+P  
 Sbjct: 658 NYLSYDNSDIGNYKALLSQIIAINRQDL- IEMSDQDKMFVTVVLRYLRLKQLRKQQLLELPT 716

Query: 720 ITDSLDTVTVSSDVMTRRRFPEDDPLKLSSSALTTFYNNQYKYFLQYVVLGLEEQDSIHPDM 779  
 L T +S +V+ FP+ PLKLS++ALT FYNNQY YFL+YVL L + +SIHPD  
 Sbjct: 717 SDYRLSTKPLSKEVIEVCFPKGIPKLKSATALTVPYNNQYNYFLKYVLNLNKTESIHPDS 776

Query: 780 RHHGTYLHRVFEILMKNOGIESFEEKLNSAINKTNOEDVFKSLYSEDAESRYSLEILEDI 839  
 R HG YLHRVFE LMK+ E F+ KL AI TNQE F+ +Y ++AE+ YSL ILEDI  
 Sbjct: 777 RIHQYLYHRVFERLMKDHTQEPFDNKLKQAIYHTNQESFFQOVYQDNAAEYSLAILEDI 836

Query: 840 ARATATILRQDSQMTVESEEEERFELMIDNTIKINGIIDRIDRLSDGSLGVVDYKSSAQKF 899  
 R+TA IL+ + + V +E+ F+L + N I ++GIIDRID+LSDGSLG+VDYKSSA +F  
 Sbjct: 837 VRSTAPILQLNQNIQVIDQEKNFQLDMGNEILVHGIIDRIDQLSDGSLGIVDYKSSANQF 896

Query: 900 DIQKFYNGLSFQLVITYIDAISR--DKEVEQKPPIFGAMYLHMQEPRQDLSKIKNLDD-LV 956  
 DI FYNGLSFQL+TY+ A+ + ++ Q +FGAMYLH+Q+P+ DL K +D+ LV  
 Sbjct: 897 DIGTFYNGLSFQLMTYLAALKQIAPHDINQ---LFGAMYLHLQDPKLDLVTFKQIDNTLV 953

Query: 957 TKNHQALTYKGLFSEAEKEFLANGKYHLKDSLYSETEIAILQAHNQSLYKKASETIKSGK 1016  
 ++ALTYKG+FSE EKE L+ G Y' K++LYS E+ L +N+ LY KA++ IK G  
 Sbjct: 954 ESIYKALTYKGIFSEVEKEHLSTGAYQTKNALYSNDELETLNLYNKYLYLKAHKHKKGH 1013

Query: 1017 FLINPYTEDAKTVGDQFKSITGFEADRHMARARALYKLPAGEKRGFLTLMQQE 1071  
 FLINPYT D KTV GDQ K+IT FEAD M +AR L LPAKEK++ FLTLM++E  
 Sbjct: 1014 FLINPYTSDGKTVQGDQLKAITRFEADLDMGQARRLVTLPAKEKKECFLTLMRKE 1068

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1400

A DNA sequence (GBSx1485) was identified in *S.agalactiae* <SEQ ID 4297> which encodes the amino acid sequence <SEQ ID 4298>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.80 Transmembrane 51 - 67 ( 44 - 69)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4121(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8799> which encodes amino acid sequence <SEQ ID 8800> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10  
 McG: Discrim Score: -20.62  
 GvH: Signal Score (-7.5): -6.25  
 Possible site: 31  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -7.80 threshold: 0.0  
 INTEGRAL Likelihood = -7.80 Transmembrane 47 - 63 ( 40 - 65)  
 PERIPHERAL Likelihood = 3.34 26  
 modified ALOM score: 2.06

\*\*\* Reasoning Step: 3

-1544-

## ----- Final Results -----

bacterial membrane --- Certainty=0.4121(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli K12]  
 Identities = 138/297 (46%), Positives = 193/297 (64%), Gaps = 16/297 (5%)

10 Query: 5 MKIDDLRKSDNVEDRRSSSGGSFSSGGSGLPILQLLLLRGSKTKLVVLIILLILG--GG 62  
 M+ R+SDNVEDRR+SSGG S GG G + S K L++LI++L+ G G  
 Sbjct: 1 MRWQGRRESNVEDRRNSSGGP-SMGGPGFRL-----PSGKGGLILLIVLVAGYYGV 52

15 Query: 63 GLTSIFNDSSSPSSYQSQNVSRSDNSATREQIDFVNKVLGSTEDFWSQEFQTQGFQNYK 122  
 LT + ++++S + D +A F + +L +TED W Q+F+ G Y+  
 Sbjct: 53 DLTGLMTGQPVVSQQQSTRSISPNEDEAAK----FTSVILATTEDTWGQQFEKMG-KTYQ 106

20 Query: 123 EPKLVLYTNSIQTCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYVIA 182  
 +PKLV+Y +TGCG G+S GPFYC AD +Y+D+SFY+++ K GA GDFA YVIA  
 Sbjct: 107 QPKLVMYRGMTRTGCAGQSGIMGPFYCPADGTVYIDLFSFYDDMKDKLGADGDFAGQGVIA 166

25 Query: 183 HEVGHHIQTELGIMDKYNRMHGLTKKEANALNVRLELQADYYAGVWAHYIRGKNLLEQG 242  
 HEVGHH+Q LGI K +++ T+ E N L+VR+ELQAD +AGVW H ++ + +LE G  
 Sbjct: 167 HEVGHHVQKLLGIEPKVRQLQONATQAEVNRLSVMELQADCFAGVWGHSMQQQGVLETG 226

Query: 243 DFEEAMNAAHAVGDDTLQKETYKLVPSFTHGTAEQRQRWFNKGFGQYGDIGHGDTF 299  
 D EEA+NAA A+GDD LQ+++ G++VPDSFTHGT++QR WF +GF GD +TF  
 Sbjct: 227 DLEEALNAAQAIGDRLQQSQGRVVPDSFTHGTSQQRYSWFKRGFDSDGPAQCNTF 283

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4299> which encodes the amino acid sequence <SEQ ID 4300>. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -6.42 . Transmembrane 48 - 64 ( 41 - 67)

35

## ----- Final Results -----

bacterial membrane --- Certainty=0.3569(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the databases:

>GP:AAC75528 GB:AE000334 orf, hypothetical protein [Escherichia coli]  
 Identities = 143/301 (47%), Positives = 195/301 (64%), Gaps = 21/301 (6%)

45 Query: 1 MKTDDLRESQVEDRRGQSSG-SFGGGGLGGGLLLQLLFSRGGWTKLVILLLLLVMG-- 57  
 M+ RES VEDRR S G S GG G +L +GG L++L+++LV G  
 Sbjct: 1 MRWQGRRESNVEDRRNSSGGPSMGGPGF-----RLPSGKGG----LILLIVLVAGYY 50

50 Query: 58 GGGLSGVLGGKPSSTNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWTKTFREKG 117  
 G L+G++ G+P S QS++ N D+A++ F S + A+TED W + F + G  
 Sbjct: 51 GVDLTGLMTGQPVVSQQ----QSTRSISPNEDEAAK----FTSVILATTEDTWGQQFEKMG 102

55 Query: 118 LTYHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMA 177  
 TY +P LV+Y G T+T CG GQ+ GPFYCP D VY+D+SFY+++ K GA GDFA  
 Sbjct: 103 KTYQQPKLVMYRGMTRTGCAGQSGIMGPFYCPADGTVYIDLFSFYDDMKDKLGADGDFAGQ 162

60 Query: 178 YVIAHEVGHHIQNELGIMDNYSARQGKSKAKANQLNVKLELQADYYAGAWANYVQGGQL 237  
 YVIAHEVGHH+Q LGI +Q ++A+ N+L+V++ELQAD +AG W + +Q QG+  
 Sbjct: 163 YVIAHEVGHHVQKLLGIEPKVRQLQONATQAEVNRLSVMELQADCFAGVWGHSMQQQGV 222

Query: 238 LEKGDIEEAMAAHAVGDDTLQEEETYGRTPVPSFTHGTSKQRQRWFDGRGYQYGDFFEHDTF 298  
 LE GD+EEA+ AA A+GDD LQ+++ GR VPDSFTHGTS+QR WF RG+ GD +TF  
 Sbjct: 223 LETGDLEEALNAAQAIGDRLQQSQGRVVPDSFTHGTSQQRYSWFKRGFDSDGPAQCNTF 283

-1545-

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/303 (63%), Positives = 241/303 (79%), Gaps = 5/303 (1%)

```

5  Query: 5  MKIDDLRKSDNVEDRRSSSGGSFSSGG-SGLPILQLLLLRGSKTKLVVLIILLGSGG 63
    MK DDLR+S VEDRR S GSF GG G +LQLL RG WKTKLV+L++LL++GGGG .
    Sbjct: 1  MKTDDLRESQQVEDRRGQSSGSFGGGGLGGGLLLQLLFSRGGWTKLVILLLLLMGGGG 60

    Query: 64  LTSIFN---DSSSPSSYQSQNVSRSDNSATREQIDFVNKVLGSTEDFWSQEFQTQGFN 120
    L+ +      S++ ++YQS V+R+ + A++EQ+ FV+KV STED+W++ F+ +G
10  Sbjct: 61  LSGVLGGKPSSTNNNAYQSSQVTRTNGDKASQEQVSFVSKVFASTEDYWKTFREKGL-T 119

    Query: 121 YKEPKLVLYTNSIQTCGIGESASGPFYCSADKKIYLDISFYNELSHKYGATGDFAMAYV 180
    Y +P LVLTY + QT CG G+++SGPFYC D+K+YLDISFYNELS KYGA GDFAMAYV
15  Sbjct: 120 YHKPTLVLYTGATQTACGRGQASSGPFYCPGDQKVYLDISFYNELSTKYGAKGDFAMAYV 179

    Query: 181 IAHEVGHHIQTELGIMDKYNMRHGLTKKEANALNVRLQADYYAGVWAHYIRGKNLLE 240
    IAHEVGHHIQ ELGIMD Y R G +K +AN LNV+LELQADYYAG WA+Y++G+ LLE
20  Sbjct: 180 IAHEVGHHIQNELGIMDNYASARQKSKAKANQLNVKLELQADYYAGAWANYVQGQGLLE 239

    Query: 241 QGDFEEMNAHAHVGGDTTLQKETYGLVPDSFTHGTAEQRQRWFNKGQYGDIGHGDTFS 300
    +GD EEAM AAHAVGGDTLQ+ETYG+ VPDSFTHGT++QRQRWF++G+QYGD +HGDTFS
25  Sbjct: 240 KGDIEEAMAAHAHVGGDTTLQEEETYGRTVPDSFTHGTSKQRQRWFDGRYQYGD FEHGDTFS 299

    Query: 301 VEH 303
    + +
30  Sbjct: 300 IPY 302

```

SEQ ID 8800 (GBS404) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 3; MW 62kDa).

30 GBS404-GST was purified as shown in Figure 218, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1401

A DNA sequence (GBSx1486) was identified in *S.agalactiae* <SEQ ID 4301> which encodes the amino acid sequence <SEQ ID 4302>. This protein is predicted to be phenylalanyl-tRNA synthetase beta chain (pheT). Analysis of this protein sequence reveals the following:

Possible site: 19  
>>> Seems to have no N-terminal signal sequence

```

40  ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB14823 GB:Z99118 phenylalanyl-tRNA synthetase (beta subunit)
[Bacillus subtilis]
Identities = 376/805 (46%), Positives = 523/805 (64%), Gaps = 6/805 (0%)

50  Query: 1  MLVSYKWLKELVDVD-VTTAELAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDT 59
    M VSYKWL++ VD+ + A LAEK++ GIEVEG+E EG+ +V+GH++ E P+
    Sbjct: 1  MFVSYKWLEDYVDLKGMDPAVLAEKITRAGIEVEGIEYKGEIGKGVVIGHVLEREQHPNA 60

    Query: 60  H-LHLCQVDVTGDDDELQVVCAPNVKTGINVIVAVPGARIADNYKIKKGKIRGMESLGMI 118
    L+ C VD G + Q++CGAPNV G V VA GA + N+KIKK K+RG ES GMI
55  Sbjct: 61  DKLNKCLVDIGARAPVQIICGAPNVDKGQKQVAVATVGAVLPGNFKIKKAKLRGEESNGMI 120

```

-1546-

Query: 119 CSLQELGLSEIIPKEFSDGIQILPEGAIPGDSIFS YLDLDEIIELSTIPNRADALSMR 178  
 CSLQELG+ ++ KE+++GI + P A G + L LDD I+EL +TPNRADA++M  
 Sbjct: 121 CSLQELGIESKLVAKEYARGIFVFPNDAETGSDALAALQLDDAILELGLTPNRADAMML 180

5 Query: 179 GVAHEVAAYIGKKVHFEEKNLIEEAERAADKISVVIESDKVLS-YSARIVKNVTVAPSPQ 237  
 GVA+EVA AI +V + + +E+A+D ISV IE + Y+A+I+KNVT+APSP  
 Sbjct: 181 GVAYEVAAILDTEVKLPQTDYPAASEQASDYISVKIEDQEANPLYTAKIIKNVTIAPSPL 240

10 Query: 238 WLQNKLMNAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLD 297  
 W+Q KLMNAGIRP NNVVD+TN+VLL YGQP+HAFD+D+F +V R A E ++TLD  
 Sbjct: 241 WMQTKLMNAGIRPHNNVVDITNFVLLLEYGQPLHAFDYDRFGSKEVVVRKAAENEMIVTLD 300

15 Query: 298 GEERDLIADDLVIAVNDQPVALAGVMGGQSTEIGSSSKTVVLEAAVFNGTSIRKTSGRLLN 357  
 +ER L AD LVI + A+AGVMGG +E+ +KT++LEAA FNG +RK S L  
 Sbjct: 301 DQERKLSADHLVITNGTKAQAVAGVMGGAESEVQEDTKTILLEAAVFNGQKVRKASKDLG 360

20 Query: 358 LRSESSSRFEKGINYD TVSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPVEVSTTLGY 417  
 LRSESS RFEKGI+ V A + AA ++ AGG+VL+G V E L E + +  
 Sbjct: 361 LRSESSVRFEKGI DPARVRLAAERAAQLIHL YAGGEVL AGTVEEDHLTIEANNIHVSADK 420

25 Query: 418 VNTRLGTELTYTDIEEVFEKLGFAISGSEVKFTVLVPRRRWDIAIQADLVEEIARIYGYE 477  
 V++ LG ++ ++ ++++LGF + ++ V VP RR DI I+ DL+EE AR+YGY+  
 Sbjct: 421 VSSVLGLTISKEELISYKRLGFTVGEADDLLVVTVPSSRRGITIEEDLIEEAARLYGYD 480

30 Query: 478 KLPTTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPKAVQFSTQATNIT 537  
 +P+TLPE T G LT Q RR+VR EGAGLS+ ITY+LT +KA F+ + + T  
 Sbjct: 481 NIPSTLPETAGTTGGLTPYQAKRRKVRRFLEGAGLSQAITYSLTNEKKATAFAIEKSLNT 540

35 Query: 538 ELMWPMFTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVYEIGKVFEQTGNPKEDLPTEVE 597  
 L PM+ +RS LR ++V +LD+++YN+AR+ ++A+YE+G VF ++ P E E  
 Sbjct: 541 VLALPMSEERSILRHS LVPNLLDSVSYNLRQTDSVALYEVGVSF--LTKEEDTKPVETE 598

40 Query: 598 TFTFALTGLVEEKDFQTKSKPVDFFYAKGIVEALFIK LK-LDVT FVAQKGLASMPGRTA 656  
 A+TGL ++ +Q + KPVDFF KGIVE L KL LD Q +HPGRTA  
 Sbjct: 599 RVAGAVTGLWRKQLWQGEKKPVDFFVVKGIVEGLLDKLNVLDSIEFVQSERKQLHPGRTA 658

45 Query: 657 TILLDGKEIGFVGQVHPQTAKQYDIPETYVAEINLSTIESQMNALIFEDITKYPVSVD 716  
 ILL+G IGF+GOVHP K+ DI ETYV E++L + + L++ I KYPVS+RD  
 Sbjct: 659 NILLNGSLIGFIGQVHPSLEKELDIKETVVFELDLHALLAETAPLVYTAIPKYPVSVD 718

50 Query: 717 IALLLAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNL 776  
 IAL+ ++V+ + S I+ +G K L + +FDVY G ++ EG KS+A+SL + NP L  
 Sbjct: 719 IALVTDKTVTSGQLESVIKEAGGKLLKEVTVFDVYEGEHMERGKKSVAFSLQYVNPQTL 778

55 Query: 777 TDEEVAKYMEKITKSLVEKVNAEIR 801  
 T+EEV K K+ K+L + A +R  
 Sbjct: 779 TEEEVTKAHSKVLKALEDTYQAVLR 803

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4303> which encodes the amino acid sequence <SEQ ID 4304>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.1283(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

60 Identities = 595/801 (74%), Positives = 687/801 (85%)

Query: 1 MLVSYKWLKELVDVDVTTAELAEKMSTTGIEVEGVETPAEGLSKLVVGHIVSCEDVPDTH 60  
 MLVSYKWLKELVD+DVT A LAEKMSTTGIEVEG+E PA+GLSKLVVGH++SCEDVP+TH  
 Sbjct: 6 MLVSYKWLKELVDIDVTPAALAEKMSTTGIEVEGIEVPADGLSKLVVGHVLSCEDVPETH 65

-1547-

Query: 61 LHLQVDTGDDELQVVCAPNVKGINVIVAVPGARIADNYKIKKGGKIRGMESLGMICS 120  
 LHLQVDTGD+ RQ+VCGAPNVK GI VIVAVPGARIADNYKIKKGGKIRGMESLGMICS  
 Sbjct: 66 LHLQVDTGDETPRQIVCGAPNVKAGIKVIVAVPGARIADNYKIKKGGKIRGMESLGMICS 125

5 Query: 121 LQELGLSESIIPKEFSDGIQILPEGAIPGDSIFS YLDLDD EIIELSITPNRADALSMRGV 180  
 LQELGLS+SIIPKEFSDGIQILPE A+PGD+IF YLDLDD IIELSITPNRADALSMRGV  
 Sbjct: 126 LQELGLSDSIIPKEFSDGIQILPEEAVPGDAIFKYLDLDDHIIELSITPNRADALSMRGV 185

10 Query: 181 AHEVAAIYGGKVVHFEENLIEEAERAADKISVVIESDKVLSYSARIVKNVTVPSPQWLQ 240  
 AHEVAAIYGGK V F +KNL E + ++ I V I SD VL+Y++R+V+NV V PSPQWLQ  
 Sbjct: 186 AHEVAAIYGKSVSFPQKNLQESDKATSEAEVAIASDNVLTYSRVVENVKVKPSPQWLQ 245

15 Query: 241 NKLMNAGIRPINNVVDVTNYVLLTYGQPMHAFDFDKFDGTTIVARNAENGEKLITLDGEE 300  
 N LMNAGIRPINNVVDVTNYVLL +GQPMHAFD+DKF+ IVAR A GE L+TLDGE+  
 Sbjct: 246 NLLMNAGIRPINNVVDVTNYVLLYFGQPMHAFDYDKFEDHKIVARAARQGESLVTLDGEK 305

20 Query: 301 RDLIADLLVIAVNDQPVVALAGVMGGQSTEIGSSSKTVVLEAAVFNGTSIRKTSGRNLNLS 360  
 RDL +DLVI V D+PVALAGVMGGQ+TEI ++S+TVVLEAAVF+G SIRKTSGRNLNLS  
 Sbjct: 306 RDLTTEDLVITVADKPVALAGVMGGQATEIDANSQTVVLEAAVFDGKSIRKTSGRNLNLS 365

25 Query: 361 ESSSRFEKGINYDVTSEAMDFAAAMLQELAGGQVLSGQVTEGVLPTEPVEVSTTLGYVNT 420  
 ESSSRFEK+NY TV EA+DFAAAMLQELA GQVLSG V G LPTEPVEVST+L YVN  
 Sbjct: 366 ESSSRFEKGVNYATVLEALDFAAAMLQELAGQVLSGHVQAGQLPTEPVEVSTSLDYVNV 425

30 Query: 421 RLGTETLYTDIEEVFEKLGFAISGSEVKFTVLVPRRRWDIAIQADLVEEIIARIYGYEKL 480  
 RLGTETL+ DI+ +F++LGF ++G E FTV VPRRRWD++I ADLVEEIIARIYGY+KLP  
 Sbjct: 426 RLGTETLFADIQRIFDQLGFLTGDETSFTVAVPRRRWDVSI PADLVEEIIARIYGYDKLP 485

35 Query: 481 TTLPEAGATAGELTSMQRLRRRVRTVAEGAGLSEIITYALTTPPEKAVQFSTQATNITELM 540  
 TTLPEAG TA ELT Q LRR+VR +AEG GL+EII+YALTTPPEKAV+F+ +++TELM  
 Sbjct: 486 TTLPEAGGTAAELTPTQALRRKVRGLAEGGLTEIISYALTTPPEKAVEFAVAPSHLTEL 545

40 Query: 541 WPMTVDRSALRQNVVSGMLDTIAYNVARKNSNLAVYEIGKVFEQTGNPKEDLPTEVETFT 600  
 WPM+V+RSALRQNV+VSGMLDT+AYNVARK SNLA+YEIGK+FEQ NPKEDLP EV F  
 Sbjct: 546 WPMSVERSALRQNMVSGMLDTVAYNVARKQSNLALYEIGKIFEQANPKEDLPNEVNHFA 605

45 Query: 601 FALTGLVEEKDFQTKSKPVDFFYAKGIVEALFIKLDVTFVAQKGLASMHGRTATILL 660  
 FA+ GLV +KDFQT+++ VDF++AKG ++ LF L L V +V K LA+MHGRTA ILL  
 Sbjct: 606 FAICGLVAQKDFQTAQAVDFYHAKGNLDTLFANLNLKVQYVPTKDLANMHGRTALILL 665

50 Query: 661 DGKEIGFVGQVHPQTAKQYDIPETYVAEINLSTIESQMNAALIFEDITKYPSVSRDIAL 720  
 D + IGFVGQVHP TAK Y IPETYVAE++++ +E+ + F +ITK+P+++RD+ALL  
 Sbjct: 666 DEQVIGFVGQVHPGTAKAYSIPETYVAELDMAALEALPSDQTFAEITKFPAMTRDVALL 725

55 Query: 721 LAESVSHHDIVSAIETSGVKRLTAIKLFDVYAGNNIAEGYKSMAYSLTFQNPNDNLTD 780  
 L VSH IV+AIE++GVKRLT+IKLFDVY G I G KSMAYSLTFQNPNDNLTD  
 Sbjct: 726 LDREVSHQAIVTAIESAGVKRLTSLIKLFDVYEGATIAGKKSMAYSLTFQNPNDNLTD 785

Query: 781 VAKYMEKITKSLVEKVNAEIR 801  
 VAKYMEKITK+L E+V AE+R  
 Sbjct: 786 VAKYMEKITKALTEQVGAEVR 806

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 55 Example 1402

A DNA sequence (GBSx1487) was identified in *S.agalactiae* <SEQ ID 4305> which encodes the amino acid sequence <SEQ ID 4306>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

60

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0653(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1548-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9769> which encodes amino acid sequence <SEQ ID 9770> was also identified.

5 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15205 GB:Z99120 transcriptional regulator [Bacillus subtilis]  
Identities = 60/169 (35%), Positives = 100/169 (58%)

10 Query: 17 ITFKVKGLDNVNILQNIATETFRQTFSDHNSSEQLQAFNESYILPVLKSEITHAESDTY 76  
+ KK +++ LQ ++IETF TF NS E ++A+ ++ L+ E+++ S +  
Sbjct: 3 VKMKKCSREDLQTLQQLSIETFNDFKEQNSPENMKAYLESFAFNTQLEKELSNMSSQFF 62

15 Query: 77 FVYLDTDLVGYLKVNWGSQQTEKDLDAFEIQRIYLLDAYQGGIGKATFEFALDLAYKS 136  
F+Y D ++ GY+KVN Q+E+ ++ EI+RIY+ +++Q G+GK A+++A +  
Sbjct: 63 FIYFDHEIAGYVKVNIDDAQSEEMGAESLEIERIYIKNSFQKHGLGKHLNKAIBIALER 122

20 Query: 137 GLDWAUWLVGVEFNHKAQAFYAKYGFEKFEHQFSVGDKVDTDWLLRKSL 185  
WLGWVE N A AFY K GF + H F +GD+ TD ++ K+L  
Sbjct: 123 NKKNIWLVGWEKNENALAFYKKMGFVQTGAHSFYMGDEEQTDLIMAKTL 171

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1403

25 A DNA sequence (GBSx1488) was identified in *S.agalactiae* <SEQ ID 4307> which encodes the amino acid sequence <SEQ ID 4308>. This protein is predicted to be phenylalanyl-tRNA synthetase (alpha subunit) (pheS). Analysis of this protein sequence reveals the following:

Possible site: 45

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3937 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

35 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9339> which encodes amino acid sequence <SEQ ID 9340> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14824 GB:Z99118 phenylalanyl-tRNA synthetase (alpha subunit)

40 [Bacillus subtilis]

Identities = 209/338 (61%), Positives = 270/338 (79%), Gaps = 2/338 (0%)

45 Query: 1 MKISTQEKLEKEM-TGNHTKELQDLRVQVLGKKGSLTELLKGLKDLNDLRPVVGKQVNEV 59  
+K QE L+++ + K + D+RVQ LGKKG +TE+L+G+ LS + RP +G NEV  
Sbjct: 5 LKQLEQEALEQVEAASSLKVVNDIRVQYLGKKGPITEVLRGMGKLSAEERPKMGALANEV 64

50 Query: 60 RDILTAFEEQAKVVEAAKIQAESESVDTLPGRQMTLGHHRVLTQTSEEIEDIFLGM 119  
R+ + A ++ + +E +++ +L +++DVTLPG + +G RH LT EEIED+F+GM  
Sbjct: 65 RERIANAIADKNEKLEEEEMKQKLAGQTIDVTLPGNPVAVGGRHPLTVVIEEIEDLFIGM 124

Query: 120 GFQVVDGFEVEKDYNNFERMNLPKDHPARDMQDTFYITEEILLRTHTSPPVQARTMDQHDHF 179  
G+ V +G EVE DYNNFE +NLPK+HPARDMQD+FYITEE L+RT TSPVQ RTM++H+  
Sbjct: 125 GYTVEEGPEVETDYNNFESLNLPEKHPARDMQDSFYITEETLMRTQTSPVQTRIMEKHE- 183

55 Query: 180 SKGPLKMISPGRVFRDRTDDATHSHQFHQIEGLVVGGENISMGDLKGTLLQLISQKMFGAER 239

-1549-

KGP+K+I PG+V+RRD DDATHSHQF QIEGLVV +NISM DLKGT+L+++KMFG +R  
 Sbjct: 184 GKGPVKIICPGKVYRRDNDATHSHQFMQIEGLVVDKNISMSDLKGTLELVAKKMFGQDR 243

Query: 240 KIRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSE 299  
 +IRLRPS+FPFTEPSVEVDV+CFKCGG GC+VCK TGWIEILGAGMVHP+VL+M+G D +  
 Sbjct: 244 EIRLRPSFFPFTEPSVEVDVTCFKCGNGCSVCKGTGWIEILGAGMVHPNVLKMAGFDPK 303

Query: 300 KYSGFAGLQGERIAMLRYGINDIRGFYQGDVRFQDQF 337  
 +Y GFAG+G ERIAML+YGI+DIR FY DVRF QF  
 Sbjct: 304 EYQGFAGMGVERIAMLRYGIDDIRHFYTNDVRFISQF 341

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4309> which encodes the amino acid sequence <SEQ ID 4310>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2806(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 305/337 (90%), Positives = 327/337 (96%)

Query: 1 MKISTQEKLEMTGNHTKELQDLRVQLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR 60  
 +K T E L+ +TGNHTKELQDLRV VLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR  
 Sbjct: 36 LKTKTLETQLSITGNHTKELQDLRVAVLGKKGSLTELLKGLKDLSNDLRPVVGKQVNEVR 95

Query: 61 DILTAFEEQAKVVEAAKIQALESESVDTVLPGRQMTLGHRHVLQTSEEIEDIFLGMG 120  
 D+LTKAFEEQAK+VEAAKIQAL++ES+DVTLPGRQMTLGHRHVLQTSEEIEDIFLGMG  
 Sbjct: 96 DLLTKAFEEQAKIVEAAKIQALDAESIDVTLPGRQMTLGHRHVLQTSEEIEDIFLGMG 155

Query: 121 FQVVDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSVPVQARTMDQHDFS 180  
 FQ+VDGFEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSVPVQART+DQHDFS  
 Sbjct: 156 FQIVDGFVEVEKDYYNFERMNLPKDHPARDMQDTFYITEEILLRTHTSVPVQARTLDQHDFS 215

Query: 181 KGPLKMISPGRVFRRDTHSHQFHQIEGLVVGGENISMGDLKGTQLQLISQKMFGAERK 240  
 KGPLKM+SPGRVFRRDTHSHQFHQIEGLVVG+NISMGDLKGT+L++I +KMFG ER  
 Sbjct: 216 KGPLKMVSPGRVFRRDTHSHQFHQIEGLVVGKNISMGDLKGTLEMIKKMFGDERS 275

Query: 241 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKQTGWIEILGAGMVHPSVLEMSGIDSEK 300  
 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCK+TGWIEILGAGMVHPSVLEMSG+D+++  
 Sbjct: 276 IRLRPSYFPFTEPSVEVDVSCFKCGGKGCNVCKKTGWIEILGAGMVHPSVLEMSGVDAKE 335

Query: 301 YSGFAGLQGERIAMLRYGINDIRGFYQGDVRFQDQF 337  
 YSGFAGLQGERIAMLRYGINDIRGFYQGD RF++QF  
 Sbjct: 336 YSGFAGLQGERIAMLRYGINDIRGFYQGDQRFSEQF 372

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1404

A DNA sequence (GBSx1489) was identified in *S.agalactiae* <SEQ ID 4311> which encodes the amino acid sequence <SEQ ID 4312>. Analysis of this protein sequence reveals the following:

Possible site: 13  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2834(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1550-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1405

A DNA sequence (GBSx1490) was identified in *S.agalactiae* <SEQ ID 4313> which encodes the amino acid sequence <SEQ ID 4314>. Analysis of this protein sequence reveals the following:

Possible site: 18

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2762(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1406

A DNA sequence (GBSx1491) was identified in *S.agalactiae* <SEQ ID 4315> which encodes the amino acid sequence <SEQ ID 4316>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8801> which encodes amino acid sequence <SEQ ID 8802> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1 Crend: 5

McG: Discrim Score: 10.13

GvH: Signal Score (-7.5): -5.07

Possible site: 23

>>> Seems to have an uncleavable N-term signal seq

ALOM program count: 1 value: -6.79 threshold: 0.0

INTEGRAL Likelihood = -6.79 Transmembrane 8 - 24 ( 6 - 27)

PERIPHERAL Likelihood = 6.26 258

modified ALOM score: 1.86

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3718(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1551-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]  
Identities = 154/232 (66%), Positives = 180/232 (77%), Gaps = 1/232 (0%)

5 Query: 41 KNVSGTPSRELSSESVLTSNVKKQLGTNIAWNQSGAFIINQNKTDLNAKVSSAPYAINI 100  
K S PS+ L+ESVLT VK Q+ ++ WN SGAFI+N NKT+L+AKVSS PYA N+ K  
Sbjct: 43 KQASEAPSQALAESVLTDAVKSQIKGSLEWNGSGAFIVNGNKTNLDAKVSSKPYADNKT 102

10 Query: 101 KVNQIVPTKANALLTKATROQYRNREETGNGRTYWKPGWHQINGLKGSYNHAVDRGHLI 160  
V + VPT ANALL+KATROQY+NR+ETGNG T W P GWHQ+ LKGSY HAVDRGHL+  
Sbjct: 103 TVGKETVPTVANALLSKATROQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGHLL 162

15 Query: 161 GYALVGSRLRGFDASTSNPKNIATQAAWANQANSNQSTGQNYETLVRKALDRHKTVRVYRV 220  
GYAL+G L GFDASTSNPKNIA Q AWANQA + STGQNYE+ VRKALD++K VRVYRV  
Sbjct: 163 GYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVYRV 222

20 Query: 221 TLIY-DRDNLSSGSHIEAKSSDGSLEFNVFIPNVQSGLLFDYATGKVKQTK 271  
TL Y ++L+ S S IEAKSSDG LEFNV +PNVQ GL DY TG+V T+  
Sbjct: 223 TLYYASNEDLVPSASQIEAKSSDGELEFNVLPNVQKGLQLDYRTGEVITVTQ 274

There is also homology to SEQ IDs 368 and 1302.

SEQ ID 8802 (GBS285) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 56 (lane 6; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 7; MW 57.5kDa).

25 GBS285-GST was purified as shown in Figure 208 (lane 7) and Figure 225 (lane 8).

GBS658 was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 134 (lane 8 & 9; MW 27kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 30 Example 1407

A DNA sequence (GBSx1492) was identified in *Sagalactiae* <SEQ ID 4317> which encodes the amino acid sequence <SEQ ID 4318>. Analysis of this protein sequence reveals the following:

Possible site: 27  
>>> Seems to have a cleavable N-term signal seq.

35 ----- Final Results -----  
bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 45 Example 1408

A DNA sequence (GBSx1493) was identified in *Sagalactiae* <SEQ ID 4319> which encodes the amino acid sequence <SEQ ID 4320>. This protein is predicted to be UDP-N-acetylglucosamine (murA). Analysis of this protein sequence reveals the following:

-1552-

Possible site: 43

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.1814(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9767> which encodes amino acid sequence <SEQ ID 9768>  
 10 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine

1-carboxyvinyltransferase [Bacillus subtilis]

Identities = 248/423 (58%), Positives = 323/423 (75%), Gaps = 5/423 (1%)

15

Query: 5   MDKIIIVEGGQTQLQGQVVEGAKNAVLPLLAATILPSQGKTLTNPILSDVFTMNNVVR 64  
           M+KIIV GGQ +L G V +EGAKNAVL++AA++L S+ K+++ +VP LSDV+T+N V+R  
 Sbjct: 1   MEKIIIVRGGQ-KLNGTVKVEGAKNAVLPVIAASLLASEEKSVICDVPTLSDVYTINEVLR 59

20

Query: 65   GLDIQVDFNCCKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPPGC 124  
           L   V F +   E+ V+AS +   AP+E+V +MRAS++V+GP+LAR GHA+V++PGGC  
 Sbjct: 60   HLGADVHF--ENNEVTVNASYALQTEAPFEYVRKMRASVLVMGPLLARTGHARVALPGGC 117

25

Query: 125   TIGSRPIDLHLKGLEAMGATITQNGGDITAE-KLKGANIYMDFPSVGATQNLMAATL 183  
           IGSRPID HLKG EAMGA I    G I A+ + +L+GA IY+DFPSVGAT+NL+MAA L  
 Sbjct: 118   AIGSRPIDQHLKGFEAMGAIEIKVGNGFIEAEVKGRLQGAKIYLDFPSVGATENLIMAAAL 177

30

Query: 184   ASGTTTIEAAREPEIVDLAQLLNKMGAKVKGAGTETLTIIGVDALHGTEHDVVQDRIEA 243  
           A GTTT+EN A+EPEIVDLA +N MG K++GAGT T+ I GV+ LHG +H ++ DRIEA  
 Sbjct: 178   AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVEKLHGVKHHIIPDRIEA 237

35

Query: 244   GTFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGEVSEEDGIRVKADTKKLKPVTVK 303  
           GTFMVAAA+T GNVLVK A+ EH   LI+K+ EMGV + +E +G+RV   K+LKP+ +K  
 Sbjct: 238   GTFMVAAAITEGNVLVKAVPEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

40

Query: 364   GGRALQGA PVMSTDLRASAALILAGMVAQGQTVVGQLTHLDRGYQFHEKLAALGANIKR 423  
           G   LQGA V +TDLRA AALILAG+VA+G T V +L HLDRGY FH+KLAALGA+I+R  
 Sbjct: 357   GPVQLQGA EVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

45

Query: 424   VSE 426  
           V++  
 Sbjct: 417   VND 419

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4321> which encodes the amino acid  
 sequence <SEQ ID 4322>. Analysis of this protein sequence reveals the following:

50

Possible site: 39

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL   Likelihood = -3.03   Transmembrane 377 - 393 ( 376 - 394)

----- Final Results -----

55

bacterial membrane --- Certainty=0.2211(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

60

&gt;GP:CAB15693 GB:Z99122 UDP-N-acetylglucosamine

1-carboxyvinyltransferase [Bacillus subtilis]

Identities = 248/423 (58%), Positives = 318/423 (74%), Gaps = 5/423 (1%)

-1553-

5 Query: 1 VDKIIIEGGQTRLEGEVVEGAKNAVLPLLAASILPSKGKTILRNVPILSDVFTMNNVVR 60  
 ++KII+ GGQ +L G V +EGAKNAVL++AAS+L S+ K+++ +VP LSDV+T+N V+R  
 Sbjct: 1 MEKIIIVRGQ-KLNGTVKVEGAKNAVLPIAASLLASEEKSVCIDVPTLSDVYTTINEVL 59

10 Query: 61 GLDIRVDFNEAANEITVDASGHILDEAPYEVVSQMRASIVVLGPILARNGHAKVSMPPGC 120  
 L V F NE+TV+AS + EAP+EYV +MRAS++V+GP+LAR GHA+V++PPGC  
 Sbjct: 60 HLGADVHFEN--NEVTVNASYALQTEAPFEYVRKMRASVLMGPLLARTGHARVALPPGC 117

15 Query: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQAD-RLQGAMIYMDFPSVGATQNLMMATL 179  
 IGSRP+ HLKG EAMGA I G I A+ RLQGA IY+DFPSVGAT+NL+MAA L  
 Sbjct: 118 AIGSRPIDQHLKGFEAMGAIEIKVGNFIEAEVKRLQGAKIYLDFPSVGATENLIMAAAL 177

20 Query: 180 ADGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEA 239  
 A+G TT+EN A+EPEIVDLA ++N MG +IRGAGT T+ I GV L GV+H ++ DRIEA  
 Sbjct: 178 AEGTTTLENVAKEPEIVDLANYINGMGGKIRGAGTGTIKIEGVKLGHVKHIIIPDRIEA 237

25 Query: 240 GTFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIQVANTPKLKPVTVK 299  
 GTFMVAAA+T GNVL++ AV EH LI+K+ EMGV++ +E G+RV +LKP+ +K  
 Sbjct: 238 GTFMVAAAITEGNVLVKGAVPEHLTSLIAKMEEMGVTIKDEGEGLRV-IGPKELKPIDIK 296

30 Query: 300 TLPHPGFPTDMQAQFTALMAVNGESTMIVETVFENRFQHLEEMRRMGLQSEILRETAMIH 359  
 T+PHPGFPTDMQ+Q AL+ +G S + ETVFENRF H EE RRM +I + +I+  
 Sbjct: 297 TMPHPGFPTDMQSOMMALLLRASGTSMTITVFENRFMHAEFFRRMNGDIKIEGRSVIIN 356

35 Query: 360 GGRQLQGAPVMSTD LRASAALILTGI VAQGVTTIVNNLVHLDRGYYQFHEKLAKLGATISR 419  
 G QLQGA V +TDLRA AALIL G+VA+G T V L HLDRGY FH+KLA LGA I R  
 Sbjct: 357 GPVQLQGA EVAATDLRAGAALILAGLVAEGHTRVTELKHLDRGYVDFHQKLAALGADIER 416

40 Query: 420 SSE 422  
 ++  
 Sbjct: 417 VND 419

An alignment of the GAS and GBS proteins is shown below.

35 Identities = 363/422 (86%), Positives = 391/422 (92%)

40 Query: 5 MDKIIIEGGQTQLQGQVVEGAKNAVLPLLAATILPSQKKTLLTNVPILSDVFTMNNVVR 64  
 +DKII+EGGQT+L+G+VVEGAKNAVLPLLA+ILPS+GKT+L NVPILSDVFTMNNVVR  
 Sbjct: 1 VDKIIIEGGQTRLEGEVVEGAKNAVLPLLAASILPSKGKTILRNVPILSDVFTMNNVVR 60

45 Query: 65 GLDIQVDFNCCKEILVDASGDILDVAPYEFVSQMRASIVVLGPILARNGHAKVSMPPGC 124  
 GLDI+VDFN EI VDASG ILD APYE+VSQMRASIVVLGPILARNGHAKVSMPPGC  
 Sbjct: 61 GLDIRVDFNEAANEITVDASGHILDEAPYEVVSQMRASIVVLGPILARNGHAKVSMPPGC 120

50 Query: 125 TIGSRPIDLHLKGLEAMGATITQNGGDITAQAELKKGANIYMDFPSVGATQNLMMATLA 184  
 TIGSRPI+LHLKGLEAMGATITQ GGDITAQA++L+GA IYMDFPSVGATQNLMMATLA  
 Sbjct: 121 TIGSRPINLHLKGLEAMGATITQKGGDITAQADRLQGAMIYMDFPSVGATQNLMMATLA 180

55 Query: 185 SGTTTIENAAREPEIVDLAQLLNKMGAQVKGAGTETLTIIIGVDALHGTEHDVVQDRIEAG 244  
 G TTIENAAREPEIVDLAQ LNMGA+++GAGTETLTI GV L G EHDVVQDRIEAG  
 Sbjct: 181 DGVTTIENAAREPEIVDLAQFLNKMGARIRGAGTETLTITGVTHLRGVEHDVVQDRIEAG 240

60 Query: 245 TFMVAAAMTSGNVLVKDAIWEHNRPLISKLMEMGVSEVEDGIRVKADTKLKPVTVKT 304  
 TFMVAAAMTSGNVL++DA+WEHNRPLISKLMEMGV V+EEE GIRV+A+T KLKPVTVKT  
 Sbjct: 241 TFMVAAAMTSGNVLIRDAVWEHNRPLISKLMEMGVSVTEEEYGIQVANTPKLKPVTVKT 300

65 Query: 305 LPHPGFPTDMQAQFTALMAVNGESTMIETVFENRFQHLEEMRRMGLQTEILRDTAMIHG 364  
 LPHPGFPTDMQAQFTALMAVNGESTM+ETVFENRFQHLEEMRRMGLQ+EILR+TAMIHG  
 Sbjct: 301 LPHPGFPTDMQAQFTALMAVNGESTMIVETVFENRFQHLEEMRRMGLQSEILRETAMIHG 360

Query: 365 GRALQGAPVMSTD LRASAALILAGMVAQGGTVVGQLTHLDRGYYQFHEKLAALGANIKRVSE 426  
 GR LQGAPVMSTD LRASAALIL G+VAQG T+V L HLDRGYYQFHEKLA LGA I RSSE  
 Sbjct: 361 GRQLQGAPVMSTD LRASAALILTGI VAQGVTTIVNNLVHLDRGYYQFHEKLAKLGATISRSE 422

65 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1554-

**Example 1409**

A DNA sequence (GBSx1494) was identified in *S.agalactiae* <SEQ ID 4323> which encodes the amino acid sequence <SEQ ID 4324>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2096(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23756 GB:AB009314 proton-translocating ATPase, epsilon  
subunit [Streptococcus bovis]

Identities = 102/138 (73%), Positives = 121/138 (86%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTPDGIRYDHASLITVTRPDGEMGILPGHINLIAPLNHQMKNRSHQEG- 59  
M +TVQVVTPDGIRYDHHA+ I+V+TPDGEMGILP HINLIAPL VH+MKI+R+

Sbjct: 1 MTFMTVQVVTPDGIRYDHANFISVKTPDGEMGILPEHINLIAPLTVHEMKIHRTDDPNH 60

Query: 60 VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTONIDEMR 119  
VDWVA+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER LE+AQ+T +IDE+R

Sbjct: 61 VDWVAINGGIIEIKDNLVTIVADSAERERDIDVSRAERAKIRAERKLEQAQSTHDIDEVR 120

Query: 120 RAEVALRRRAINRISVGKK 137

RA+VALRRA+NRISVG K

Sbjct: 121 RAQVALRRALNRISVGNK 138

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4325> which encodes the amino acid sequence <SEQ ID 4326>. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence.

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2539(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 100/138 (72%), Positives = 119/138 (85%), Gaps = 1/138 (0%)

Query: 1 MAQLTVQVVTPDGIRYDHASLITVTRPDGEMGILPGHINLIAPLNHQMKNRSHQ-EG 59

M Q+TVQVVTPDGI+YDHHA I+V TPDGEMGILP HINLIAPL VH+MKI R + E

Sbjct: 1 MTQMTVQVVTPDGIKYDHHAKFISVTTTPDGEMGILPNHINLIAPLQVHEMKIRRGGEDEK 60

Query: 60 VDWVAVNGGIIEVNEDQVTIVADSAERARDIDLNRAERAKERAERALEKAQTTONIDEMR 119

VDW+A+NGGIIE+ ++ VTIVADSAER RDID++RAERAK RAER + +A+TT NIDE+R

Sbjct: 61 VDWIAINGGIIEIKDNVVTIVADSAERDRDIDVSRAERAKLRAEREIAQAETTHNIDEVR 120

Query: 120 RAEVALRRRAINRISVGKK 137

RA+VALRRA+NRI+V KK

Sbjct: 121 RAKVALRRALNRINVSKK 138

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1555-

**Example 1410**

A DNA sequence (GBSx1495) was identified in *S.agalactiae* <SEQ ID 4327> which encodes the amino acid sequence <SEQ ID 4328>. Analysis of this protein sequence reveals the following:

```

Possible site: 60
5  >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to the beta subunit of the S.mutans ATPase:

>GP:AAD13383 GB:U31170 ATPase, beta subunit [Streptococcus mutans]
Identities = 435/466 (93%), Positives = 455/466 (97%)

15 Query: 1  MSSGKIAQVVGPPVDVVFASGDKLPEINNALLIVYKNGDKSQKVVLEVALELGDGLVRTIA 60
    MS+GKIAQVVGPPVDV FA+ DKLPEINNALL+VYK+GDKSQ++VLEVALELGDGLVRTIA
    Sbjct: 1  MSTGKIAQVVGPPVDVAFATDDKLPEINNALLVYKDGDKSQRIVLEVALELGDGLVRTIA 60

20 Query: 61  MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEFPAEDAERQPIHKKAP 120
    MESTDGLTRGLEV DTGRAISVPVGK+TLGRVFNVLGD IDL++PFAEDAERQPIHKKAP
    Sbjct: 61  MESTDGLTRGLEVFDTGRAISVPVGKETLGRVFNVLGDTIDLDKPFPAEDAERQPIHKKAP 120

25 Query: 121 SFDELSTSSSEILETGKIVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIHQEHGGIS 180
    SFD+LSTS+EILETGKIVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIHQEHGGIS
    Sbjct: 121 SFDDLSTSTEILETGKIVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIHQEHGGIS 180

30 Query: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNPPGARMRVALTGLTIAEYFRDVE 240
    VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNPPGARMRVALTGLTIAEYFRDVE
    Sbjct: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNPPGARMRVALTGLTIAEYFRDVE 240

35 Query: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGSVTSI 300
    GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGSVTSI
    Sbjct: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMPSAVGYQPTLATMGQLQERITSTKKGSVTSI 300

40 Query: 301 QAIYVPADDYTDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALTPEIVGDEH 360
    QAIYVPADDYTDPAAPATAFAHLDSTTNLER+LTQMGIYPAVDPLASSSRAL+PEIVG EH
    Sbjct: 301 QAIYVPADDYTDPAAPATAFAHLDSTTNLERRLTQMGIYPAVDPLASSSRALSPEIVGQEH 360

45 Query: 361 YEVADEVQRLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAETFTGQ 420
    Y+VATEVQ VLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAE FTGQ
    Sbjct: 361 YDVATEVQHVLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAEQFTGQ 420

Query: 421 PGSYVPVEETVRGFKIILDGKHDQIPEDAFRVMVGIEDVIAKAEKM 466
    PGSYVPV ETVRGFKIIL+GK+D++PEDAFR VG IEDV+ KA+KM
    Sbjct: 421 PGSYVPAETVRGFKIILEGKYDELPEDAFRSVGAIEDVVEKAKKM 466

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4329> which encodes the amino acid sequence <SEQ ID 4330>. Analysis of this protein sequence reveals the following:

```

50 Possible site: 60
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.0275(Affirmative) < succ>
55                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 440/468 (94%), Positives = 456/468 (97%)

60 Query: 1  MSSGKIAQVVGPPVDVVFASGDKLPEINNALLIVYKNGDKSQKVVLEVALELGDGLVRTIA 60

```

-1556-

Sbjct: 1 MSSGKIAQVVGPFVVDV+FASGDKLPEINNALIVYK+ DK QK+VLEVALELGDG+VRTIA 60  
 MSSGKIAQVVGPFVVDVVFASGDKLPEINNALIVYKSDKKQKIVLEVALELGDGMVRTIA 60

5 Query: 61 MESTDGLTRGLEVLDTGRAISVPVGKDTLGRVFNVLGDAIDLEEPFAEDAERQPIHKKAP 120  
 MESTDGLTRGLEVLDTGRAISVPVGK+TLGRVFNVLG+ IDLEEPFAED +RQPIHKKAP  
 Sbjct: 61 MESTDGLTRGLEVLDTGRAISVPVGKETLGRVFNVLGETIDLEEPFAEDVDVRQPIHKKAP 120

10 Query: 121 SFDELSTSSSEILETGIIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAHEGGIS 180  
 SFDELSTSSSEILETGIIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAHEGGIS  
 Sbjct: 121 SFDELSTSSSEILETGIIKVIDLLAPYLKGGKVGLFGGAGVGKTVLIQELIHNIQAHEGGIS 180

15 Query: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240  
 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE  
 Sbjct: 181 VFTGVGERTREGNDLYWEMKESGVIEKTAMVFGQMNEPPGARMRVALTGLTIAEYFRDVE 240

20 Query: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMP SAVGYQPTLATMGLQERITSTKGSVTSI 300  
 GQDVLLFIDNIFRFTQAGSEVSALLGRMP SAVGYQPTLATMGLQERITST+KGSVTSI  
 Sbjct: 241 GQDVLLFIDNIFRFTQAGSEVSALLGRMP SAVGYQPTLATMGLQERITSTQKGSVTSI 300

25 Query: 301 QAIYVPADDDYDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRAL+PEIVGDEH 360  
 QAIYVPADDDYDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRAL+PEIVG+EH  
 Sbjct: 301 QAIYVPADDDYDPAAPATAFAHLDSTTNLERKLTQMGIYPAVDPLASSSRALSPEIVGEEH 360

30 Query: 361 YEVADEVQVRLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAETFTGQ 420  
 Y VATEVQVRLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAE FTG  
 Sbjct: 361 YAVATEVQVRLQRYRELQDIIAILGMDELSDEEKTIVGRARRIQFFLSQNFNVAEQFTGL 420

Query: 421 PGSYVPVEETVRGFKELDKHDIQIPEDAFRVMGGIEDVIKAEKMN 468  
 PGSYVPV +TVRGFKEL+GK+D++PEDAFR VG IEDVI KAEK +  
 Sbjct: 421 PGSYVPVADTVRGFKELLEGKYDELPEDAFRSVGPIEDVIKAEKMGF 468

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1411

35 A DNA sequence (GBSx1496) was identified in *S. agalactiae* <SEQ ID 4331> which encodes the amino acid sequence <SEQ ID 4332>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1889(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23754 GB:AB009314 proton-translocating ATPase, gamma subunit  
 [Streptococcus bovis]  
 Identities = 252/293 (86%), Positives = 278/293 (94%), Gaps = 2/293 (0%)

50 Query: 1 MAGSLSEIKDKILSTEKTSKITSAMQMVSSAKLVKSEQAARDFQVYASKIRQITTNLLKS 60  
 MAGSLSEIK KI+ST+KTS IT AMQMV+AKL KSEQAA+DFQVYASKIRQITT+LLKS  
 Sbjct: 1 MAGSLSEIKGKIISTQKTSHTGAMQMVSAAKLTKSEQAAKDFQVYASKIRQITTDLLKS 60

55 Query: 61 DLVSGSDNPMLSSRPVKKTYIVITSDKGLVGGYNSKILKAMMDTITDYHTENDDYAIIS 120  
 +LV+GS NPML++RPVKKTYIVITSDKGLVGGYNSKILKAMMD I +YH ++ +YAI+  
 Sbjct: 61 ELVNGSKNPMLAARPVKKTYIVITSDKGLVGGYNSKILKAMMDLIEEYH-QDGNYAIIA 119

60 Query: 121 IGSVGSDFFKARGMNVFELRGLEDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180  
 IG +G+DFFKARGMNV FELRGLEDQPSF+QVG IIA++VEMYKNELFDELYVCYNHHVN  
 Sbjct: 120 IGGIGADFFKARGMNVFELRGLEDQPSFEQVGNIIAKSVEMYKNELFDELYVCYNHHVN 179

Query: 181 SLTSQVRMQMLPIKELDAEEASEDRVTGFELEPNREVILEQLLPQYTESLIYGAIIDA 240

-1557-

SLTSQVR+QQMLPI ELDA+EA+E+ V +GFELEPNRE+ILEQLLPQYTESLIYGAI+DA  
 Sbjct: 180 SLTSQVRVQQMLPIAELDADEAAEEGV-SGFELEPNREMILEQLLPQYTESLIYGAI+DA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE 293  
 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE

Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE 291

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4333> which encodes the amino acid sequence <SEQ ID 4334>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1969(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 251/293 (85%), Positives = 275/293 (93%), Gaps = 2/293 (0%)

Query: 1 MAGSLSEIKDKILSTEKTSKITSAMQMVS SAKLVKSEQAARDFQVYASKIRQITTNLLKS 60  
 MAGSLSEIK KI+STEKTSKITSAM+MVSSAKLVKSEQAARDFQ+YASKIRQITT+LLKS

Sbjct: 1 MAGSLSEIKAKIISTEKTSKITSAMRMVSSAKLVKSEQAARDFQIYASKIRQITDILLKS 60

Query: 61 DLVSGSDNPMLSSRPVKKTGYIVITSDKGLVGGYNSKILKAMMDTITDYHTENDDYAIIS 120  
 +L GSDNPML SRPVKKTGYIVITSDKGLVGGYNSKILK++MD IT+YH + DY IIS

Sbjct: 61 ELTIGSDNPMLVSRPVKKTGYIVITSDKGLVGGYNSKILKSVMDMITEYHADG-DYEIIS 119

Query: 121 IGSVGSDFFKARGMNVSFELRGLDQPSFDQVGKIIAQAVEMYKNELFDELYVCYNHHVN 180  
 IGSVGSDFFKARGMNV+FELRGL DQPSF+QV +II+Q+V+M+ NE+FDLYVCYNHHVN

Sbjct: 120 IGSVGSDFFKARGMNVAFELRGLADQPSFEQVRQIISQSDMFVNEIFDELYVCYNHHVN 179

Query: 181 SLTSQVRMQQMLPIKELDAEEASEDRVITGFELEPNREVILEQLLPQYTESLIYGAIIDA 240  
 SLTSQVR+QQMLPI +L A+EA+E+ V TGFELEPNR IL+QLLPQ+TESLIYGAIIDA

Sbjct: 180 SLTSQVRVQQMLPISDLVADEAAEEGV-TGFELEPNRHDILDQLLPQFTESLIYGAIIDA 238

Query: 241 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE 293  
 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE

Sbjct: 239 KTAEHAAGMTAMQTATDNAKNVINDLTIQYNRRARQAAITQEITEIVAGANALE 291

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1412

A DNA sequence (GBSx1497) was identified in *S.agalactiae* <SEQ ID 4335> which encodes the amino acid sequence <SEQ ID 4336>. Analysis of this protein sequence reveals the following:

Possible site: 25  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1963(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

-1558-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1413

A DNA sequence (GBSx1498) was identified in *S.agalactiae* <SEQ ID 4337> which encodes the amino acid sequence <SEQ ID 4338>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein is similar to the alpha subunit of the proton-translocating ATPase from *S.bovis*:

15       >GP:BAA23753 GB:AB009314 proton-translocating ATPase, alpha subunit  
          [Streptococcus bovis]                               Length = 501  
          Identities = 482/501 (96%), Positives = 497/501 (98%)

20       Query: 1    MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60  
                  MAINAQEISALIKKQIE+QPNFDVTETG+VTYIGDGIARARGLDNAMSGELLEFSNGA+  
      Sbjct: 1    MAINAQEISALIKKQIENFQPNFDVTETGVVITYIGDGIARARGLDNAMSGELLEFSNGAF 60

25       Query: 61   GMAQNLESNDVGIIILGDFSEIREGDVVKRTGKIMEVPVGEAMIGRVVNPLGQPV DGLGE 120  
                  GMAQNLESNDVGIIILGDFS IREGD VKRTGKIMEVPVGEA+IGRVVNPLGQPV DGLG+  
      Sbjct: 61   GMAQNLESNDVGIIILGDFSTIREGDEVKRTGKIMEVPVGEALIGRVVNPLGQPV DGLGD 120

30       Query: 121 IETTATRPVETPAPGVMQRKSVFEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180  
                  I+TTATRPVETPAPGVMQRKSV EPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI  
      Sbjct: 121 IKTTATRPVETPAPGVMQRKSVSEPLQTGLKAIDALVPIGRGQRELIIGDRQTGKTSVAI 180

35       Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240  
                  DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLL+IAPY  
      Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLYIAPY 240

40       Query: 241 AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLHSRLLER 300  
                  AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLHSRLLER  
      Sbjct: 241 AGVAMAEFMYNGKHLIVYDDLKQAVAYRELSLLLRPPGREAYPGDVFYLHSRLLER 300

      Query: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG 360  
                  SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG  
      Sbjct: 301 SAKVSDALGGGSITALPFIETQAGDISAYIATNVISITDQIFLQENLFNSGIRPAIDAG 360

      Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFGSDLDAATQAKLNRGRRTVEVL 420  
                  -----

-1559-

bacterial cytoplasm --- Certainty=0.3654 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 477/501 (95%), Positives = 490/501 (97%)

Query: 1 MAINAQEISALIKKQIEDFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60  
 +AINAQEISALIKKQIE+FQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY  
 10 Sbjct: 1 LAINAQEISALIKKQIENFQPNFDVTETGIVTYIGDGIARARGLDNAMSGELLEFSNGAY 60

Query: 61 GMAQNLESNDVGIIILGDFSEIREGDDVVKRTGKIMEVPVGEAMIGRVVNPLGQPVVDGLGE 120  
 GMAQNLESNDVGIIILGDFSEIREGDDVVKRTGKIMEVPVGEA+IGRVVNPLGQPVVDGLG+  
 15 Sbjct: 61 GMAQNLESNDVGIIILGDFSAIREGDDVVKRTGKIMEVPVGEALIGRVVNPLGQPVVDGLGD 120

Query: 121 IETTATRPVETPAPGVMQKRSVFEPLQTGLKADALVPIGRGQRELIIGDRQTGKTSVAI 180  
 IETT RVPVETPAPGVMQKRSV EPLQTGLKADALVPIGRGQRELIIGDRQTGKTSVAI  
 20 Sbjct: 121 IETTGFPRVETPAPGVMQKRSVSEPLQTGLKADALVPIGRGQRELIIGDRQTGKTSVAI 180

Query: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240  
 DAILNQKGQDMICIYVAIGQKESTVRTQVETLR+YGALDYTIVVTASASQPSPLLFIAPY  
 25 Sbjct: 181 DAILNQKGQDMICIYVAIGQKESTVRTQVETLRKYGALDYTIVVTASASQPSPLLFIAPY 240

Query: 241 AGVAMAEFMYNGKHVLIVYDDLSKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER 300  
 AGVAMAEFMY GKHVLIVYDDLSKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER  
 25 Sbjct: 241 AGVAMAEFMYQGHVLIVYDDLSKQAVAYRELSLLLRPPGREAYPGDVFYLSRLLER 300

Query: 301 SAKVSDALGGGSITLPPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360  
 SAKVSD LGGGSITLPPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG  
 30 Sbjct: 301 SAKVSDDLGGGSITLPPFIETQAGDISAYIATNVISITDGQIFLQENLFNSGIRPAIDAG 360

Query: 361 SSVSRVGGAAQIKAMKRVAGTLRLDLASYRELEAFTQFGSDLDAAATQAKLNRGRRTVEVL 420  
 SSVSRVGG+AQIKAMK+VAGTLRLDLASYRELEAFTQFGSDLDAAATQAKLNRGRRTVE+L  
 35 Sbjct: 361 SSVSRVGGSAQIKAMKKVAGTLRLDLASYRELEAFTQFGSDLDAAATQAKLNRGRRTVEIL 420

Query: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVNDILAFEEALYDYFDAHYDNLFTETIRTTKD 480  
 KQPLHKPLPVEKQVVILYALTHGFLDDVPV+DILAFEEALYDYFD HY++LFTETIRTTKD  
 35 Sbjct: 421 KQPLHKPLPVEKQVVILYALTHGFLDDVPVDDILAFEEALYDYFDVHYDNLFTETIRTTKD 480

Query: 481 LPPEAELDAAIQAFKDQSQFK 501  
 LPPEA LDAAI+AFK+ S FK  
 40 Sbjct: 481 LPPEAALDAAIKAFKEHSNFK 501

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 45 vaccines or diagnostics.

#### Example 1414

A DNA sequence (GBSx1499) was identified in *S.galactiae* <SEQ ID 4341> which encodes the amino acid sequence <SEQ ID 4342>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1896 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA23752 GB:AB009314 proton-translocating ATPase, delta subunit  
 [Streptococcus bovis]  
 60 Identities = 98/178 (55%), Positives = 127/178 (71%)

-1560-

Query: 1 MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAAALIDIFETSELEGLSSLA VSHDEKQ 60  
 M+KKTQAL+EQY+KSLVE+AIE + ++Q E AL+ +FE + L LSSL VS DEK  
 Sbjct: 1 MDKKTQALVEQYAKSLVEIAIEKDSLAELQSETEALLSVFEETNLADFLSSLVVS RDEKV 60

5 Query: 61 HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALSP 120  
 V+ LQ S S Y+ NFLEVI+QNEREA L IL+ V ++ + Q+ I +TTAVAL+  
 Sbjct: 61 KLVRLQLQESSVYMMNFLEVILQNEREAFLKAILEGVQKDFVIATNQHDIVVTTAVALTD 120

10 Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVNVNANKVIDASVRNQLHQFKMKLK 178  
 EQKER+ + K + G+LVE+ID SI+GGFV+N NNKVID S+R QL +FKM LK  
 Sbjct: 121 EQKERILALVAEKFGVKAGKLVENIDESILGGFVINNVNANKVIDTSIRRLQOEFKMNLK 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4343> which encodes the amino acid sequence <SEQ ID 4344>. Analysis of this protein sequence reveals the following:

15 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1668(Affirmative) < succ>  
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/178 (48%), Positives = 125/178 (69%)

25 Query: 1 MNKKTQALIEQYSKSLVEVAIEHKIVEKIQQEVAAALIDIFETSELEGLSSLA VSHDEKQ 60  
 M KK QALIEQY+KSLVEVA EH ++ +Q +V A+++ F T+ L+ LSS AV H EK  
 Sbjct: 1 MTKKEQALIEQYAKSLVEVASEHHS LDALQADVLATLETFTVTNLQSLSSQAVPHA EKI 60

30 Query: 61 HFVKTLQTSCSTYLVNFLEVIVQNEREALLYPILKSVDQELIKVNGQYPIQITTAVALSP 120  
 + L+ + S Y+ NFL +I+QNEREA LY +L++V E+ V+ QY + +T+++ L+  
 Sbjct: 61 KLLTLLKGNNVYMMNFNLNLILQNEREAYLYQMLQAVLNEIAIVSNQYDVTVTSSLP LTE 120

35 Query: 121 EQKERLFDIAKTKLALPNGQLVEHIDPSIVGGFVNVNANKVIDASVRNQLHQFKMKLK 178  
 EQK R+ + K A+ G+L+E +DPS++GGF+++ NNKVID S+R QL FKM LK  
 Sbjct: 121 EQKSRVRAVAKKFAVTAGRLIEKVDPSLIGGFIIISVNNKVIDTSIRRLQAFKMNLK 178

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### 40 Example 1415

A DNA sequence (GBSx1500) was identified in *S.agalactiae* <SEQ ID 4345> which encodes the amino acid sequence <SEQ ID 4346>. This protein is predicted to be ATP synthase b chain (atpF). Analysis of this protein sequence reveals the following:

45 Possible site: 33  
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----  
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]  
 Identities = 103/165 (62%), Positives = 130/165 (78%)

55 Query: 1 MSILINSTTIGDIIIVSGSVLLLLFILIKTFANKQITGIFEAREQKIANDIDTAEQARQQA 60  
 MS LIN T++G+++IV+GS +LL +L+K FAW Q+ IF+ RE+KIA DID AE +RQ A  
 Sbjct: 1 MSTLINGTSLGNLLIVTGSFILLLLLLVKKFAWSQLAAIFKTREEKIAKDIDDAENSRQNA 60

-1561-

Query: 61 EAFATKREEELSNKTEANQIIDNAKETGLAKGDQIIEAKTEADRLKEKAHQDIAQNKA 120  
 + KR+ EL+ AK EA QIIDNAKETG A+ +II+EA EA RLK+KA+QDIA +KA  
 Sbjct: 61 QVLENKRQVELNQAKDEAAQIIDNAKETGKAQESKIIIEAHEEAGRLKDKANQDIATSKA 120

Query: 121 EALADVKGGEVADLTVLLAEKIMVSNLDKEAQSNIIDSYIKKLGA 165  
 EAL+ VK +VADL+VLLAEKIM NLDK AQ +LIDSY+ KLGDA  
 Sbjct: 121 EALSSVKADVADLSVLLAEKIMAKNLDKTAQGDILIDSYLDKLGDA 165

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4347> which encodes the amino acid sequence <SEQ ID 4348>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have a cleavable N-term signal seq.

- 15 ----- Final Results -----  
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the databases:

>GP:AAD13379 GB:U31170 ATPase, b subunit [Streptococcus mutans]  
 Identities = 88/159 (55%), Positives = 122/159 (76%)

- 25 Query: 6 GELVGNFILVTGSGVIVLLLLIKKFAWGAIESILQTRSQQISRDIQAEQSRLSAQQLEAK 65  
 G +GN ++VTGS I+LLLL+KKFAW + +I +TR ++I++DID AE SR +AQ LE K  
 Sbjct: 7 GTSLGNLLIVTGSFILLLLLVKKFAWSQLAAIFKTREEKIADIDDAENSRQNAQVLENK 66
- Query: 66 SQANLDSRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAV 125  
 Q L+ ++ +A++II +AKE G+ Q K++ EA +EA RLK+KA DI SK++A+S+V  
 30 Sbjct: 67 RQVELNQAKDEAAQIIDNAKETGKAQESKIIIEAHEEAGRLKDKANQDIATSKAEALSSV 126
- Query: 126 KTEMSDLTVLLAEKIMGANLKDKAQSQLIDSYLDDLGEA 164  
 K +++DL+VLLAEKIM NLDKTAQ LIDSYLD LG+A  
 Sbjct: 127 KADVADLSVLLAEKIMAKNLDKTAQGDILIDSYLDKLGDA 165

An alignment of the GAS and GBS proteins is shown below.

Identities = 81/156 (51%), Positives = 115/156 (72%)

- 40 Query: 10 IGDIIIVSGSVLLLLFILIKTFFAWKQITGIFEAREQKIANDIDTAEQARQQAEAFATKREE 69  
 +G+ I+V+GSV++L +LIK FAW I I + R Q+I+ DID AEQ+R A+ K +  
 Sbjct: 9 VGNFILVTGSGVIVLLLLIKKFAWGAIESILQTRSQQISRDIQAEQSRLSAQQLEAKSQA 68
- Query: 70 ELSNAKTEANQIIDNAKETGLAKGDQIIEAKTEADRLKEKAHQDIAQNKAELADVKG 129  
 L ++ +A++II +AKE G +GD+++EA EA RLKEKA DI Q+K++A++ VK E  
 45 Sbjct: 69 NLDASRLQASKIISDAKEIGQLQGDKLVAEATDEAKRLKEKALTDIEQSKSDAISAVKTE 128
- Query: 130 VADLTVLLAEKIMVSNLDKEAQSNIIDSYIKKLGA 165  
 ++DLTVLLAEKIM +NLDK AQS LIDSY+ LG+A  
 Sbjct: 129 MSDLTVLLAEKIMGANLKDKAQSQLIDSYLDDLGEA 164

SEQ ID 4346 (GBS169) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 6; MW 18kDa).

The GBS169-His fusion product was purified (Figure 200, lane 11) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 250). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1562-

**Example 1416**

A DNA sequence (GBSx1501) was identified in *S.agalactiae* <SEQ ID 4349> which encodes the amino acid sequence <SEQ ID 4350>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -11.73    Transmembrane    20 - 36 ( 14 - 42)
    INTEGRAL    Likelihood = -5.20    Transmembrane    207 - 223 ( 206 - 228)
    INTEGRAL    Likelihood = -4.35    Transmembrane    78 - 94 ( 73 - 97)
10  INTEGRAL    Likelihood = -4.09    Transmembrane    113 - 129 ( 113 - 133)
    INTEGRAL    Likelihood = -2.39    Transmembrane    174 - 190 ( 174 - 190)

----- Final Results -----
    bacterial membrane --- Certainty=0.5692(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA23750 GB:AB009314 proton-translocating ATPase, a subunit
    [Streptococcus bovis]
20  Identities = 149/238 (62%), Positives = 180/238 (75%)

Query: 1  MESTSNPTVSFLGIDFDLTILAMSLLTITIIIFILVFWASRKMTIKPKGKQNVLEYVYELV 60
        ME++ NPT  GI+FDLTILAMSLLT+ I F ++FWA+RKMT+KPKGKQN +EYVYE V
Sbjct: 1  METSVNPTAHVFGIEFDLTILAMSLLTVIISFGIIFWATRKM TLKPKGKQNFIEYVYEFV 60

25  Query: 61  NNTISQNLGHYTKNYSLLMFILFSFVFIANNLGLMTSLKTHEHFWTSPTANFGVDITLS 120
        NTI  NLG YT  YSLLMF  F F+ IANNLGL+ L++ ++NFWTSPT+  VD T S
Sbjct: 61  QNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDC TWS 120

30  Query: 121 LLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVTNVA SLALRLFGNIFSGEVV 180
        L+VA + H+EG+RKKG+ YLKG+LSP P MLPMN+LE+ TNV SLALRLFGNI++GEVV
Sbjct: 121 LIVAIVVHVEGVRKKGVKAYLKGYSFPFMMLEPMNILEQFTNVLSLALRLFGNIYAGEVV 180

35  Query: 181 TGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFIILSSSYIGHKVGDEEE 238
        T L++ S P A ALN+ W AFS FIG IQAYVF ILSS YI K+ DE+E
Sbjct: 181 TALIVGFGTKSLIFAPFALALNLAWAFSAFIGCIQAYVFTILSSKYISEKLPEDEDE 238

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4351> which encodes the amino acid sequence <SEQ ID 4352>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 33
    >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -4.73    Transmembrane    79 - 95 ( 72 - 97)
    INTEGRAL    Likelihood = -4.35    Transmembrane    115 - 131 ( 112 - 132)
    INTEGRAL    Likelihood = -2.13    Transmembrane    200 - 216 ( 197 - 216)
45  ----- Final Results -----
    bacterial membrane --- Certainty=0.2890(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 124/239 (51%), Positives = 169/239 (69%), Gaps = 3/239 (1%)

55  Query: 1  MESTSNPTVSFLGIDFDLTILAMSLLTITIIIFILVFWASRKMTIKPKGKQNVLEYVYELV 60
        ME  P + I F+LT+LA+ ++TI I+F VFWASR+M +KP+GKQ LEY+ V
Sbjct: 1  MEEAKIPMLKLGPIITFNLITLLAVCIVTIAIVFAFVFWASRQMKLKPEGKQTALEYLISFV 60

    Query: 61  NNTISQNLGH-YTKNYSLLMFILFSFVFIANNLGLMTSLKT-HEHFWTSPTANFGVDIT 118
        + ++L H K+YSLL+ F +F FV +ANNLGL T L+T + +N WTSPTAN D+
60  Sbjct: 61  DGIGEEHLDHNLQKSYSLLLEFTIFLFVAVANNLGLFTKLETVNGYNLWTSPTANLAFDLA 120

```

-1563-

Query: 119 LSLLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVTNVASLALRLFGNIFSGE 178  
 LSL + + HIEG+R++G+ +LK +P P M PMNLEE TN SLA+RLFGNIF+GE  
 Sbjct: 121 LSLFTLTMVHIEGVRRRGLVAHLKRLATPWP-MTPMNLLEFTNFLSLAIRLFGNIFAGE 179

5 Query: 179 VVTGLLLQLAVLSPFTGPLAFALNIVWTAFSMFIGFIQAYVFILSSSYIGHKVHGDEE 237  
 VVTGL++QLA + P+AF +N+ WTAFS+FI IQA+VF L+++Y+G KV+ EE  
 Sbjct: 180 VVTGLIVQLANYRVYWWPIAFLVNMWTAFSVFISCIQAFVFTKLTATYLGKKVNESEE 238

A related GBS gene <SEQ ID 8803> and protein <SEQ ID 8804> were also identified. Analysis of this  
 10 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 1  
 McG: Discrim Score: -3.50  
 GvH: Signal Score (-7.5): -3.36  
 Possible site: 29

>>> Seems to have no N-terminal signal sequence

ALOM program count: 5 value: -11.73 threshold: 0.0

INTEGRAL Likelihood = -11.73 Transmembrane 20 - 36 ( 14 - 42)

INTEGRAL Likelihood = -5.20 Transmembrane 207 - 223 ( 206 - 228)

INTEGRAL Likelihood = -4.35 Transmembrane 78 - 94 ( 73 - 97)

INTEGRAL Likelihood = -4.09 Transmembrane 113 - 129 ( 113 - 133)

INTEGRAL Likelihood = -2.39 Transmembrane 174 - 190 ( 174 - 190)

PERIPHERAL Likelihood = 5.30 156

modified ALOM score: 2.85

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5692(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF01818(301 - 1014 of 1314)

GP|2662321|dbj|BAA23750.1||AB009314(1 - 238 of 239) proton-translocating ATPase, a subunit  
 {Streptococcus bovis}

%Match = 35.0

%Identity = 62.2 %Similarity = 78.6

Matches = 148 Mismatches = 51 Conservative Sub.s = 39

204 234 264 294 324 354 384 414  
 XANCQTLMLPGVGFIERYSICVYILSKIDDNLEKKEG\*GLESTSNPTVSFLGIDFDLTILAMSLTTITIFILVFWA  
 :||: ||| :||:|||||||: | | :|||  
 METSVNPTAHVFGIEFDLTILAMSLTTVIISFGIIFWA  
 10 20 30

444 474 504 534 564 594 624 654  
 SRKMTIKPKGKQNVLEYVYELVNNTISQNLGHYTKNYSLLMFILFSFVFIANNLGLMTSLKTHEHNFWTSPTANFGVDIT  
 :|||:|||||: ||||: ||| || || |||||: | :||:|||||: |::: |||||: || |  
 TRKMTLKPKGKQNFIEVYEFVQNTIKPNLGEYTPKYSLLMFTFFFFILIANNLGLLVKLESEDYNFWTSPTSTIMVDCIT  
 50 60 70 80 90 100 110

684 714 744 774 804 834 864 894  
 LSLLVAFICHIEGIRKKGIGGYLKGFLSPTPAMLPNNLLEEVTNVASLALRLFGNIFSGEVVTGLLLQLAVLSPFTGPLA  
 ||:| : |:|:||||: |||: || | ||||:|: || |||||:|: ||||| |::: | :|  
 WSLIVAIVVHVEGVRRKKGVKAYLKGYSPPFPMNLPNNILEQFTNVLSLALRLFGNIYAGEVVTALIVGFGTKSLIFAPFA  
 130 140 150 160 170 180 190

924 954 984 1014 1044 1074 1104 1134  
 FALNIVWTAFSMFIGFIQAYVFILSSSYIGHKVHGDEEE\*EKRGECQYLLIVQRLVISLSYLALCFSYLS\*LRLHHGN  
 :||: | ||| || ||||| ||| | : ||:|  
 LALNLAWVAFAFIGCIQAYVFILSSKYISEKLPEDEFT  
 210 220 230

-1564-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1417

A DNA sequence (GBSx1502) was identified in *S.agalactiae* <SEQ ID 4353> which encodes the amino acid sequence <SEQ ID 4354>. This protein is predicted to be ATP synthase c subunit (atpE). Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -4.62    Transmembrane    48 - 64 ( 42 - 65)

----- Final Results -----
                bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA23749 GB:AB009314 proton-translocating ATPase, c subunit [Streptococcus bovis]
Identities = 56/65 (86%), Positives = 59/65 (90%)
```

```
Query: 1  MNLAILALGFAVMGVSIGEGILVANIAKSAARQPEMFSLQTLMTGVAFIEGTFFVLFA 60
        +NL ILALG AV+GVS+GEGILVANIAKSAARQPEMFSLQTLMF GVAFIEGTFFVL A
Sbjct: 2  LNLKILALGLAVLGVSLGEGILVANIAKSAARQPEMFSLQTLMFLGVAFIEGTFFVLLA 61
```

```
Query: 61  FTFLV 65
        TF V
Sbjct: 62  STFFV 66
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4355> which encodes the amino acid sequence <SEQ ID 4356>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -5.26    Transmembrane    47 - 63 ( 41 - 64)

----- Final Results -----
                bacterial membrane --- Certainty=0.3102(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00920 GB:AF001955 UncE [Streptococcus sanguinis]
Identities = 50/66 (75%), Positives = 58/66 (87%), Gaps = 1/66 (1%)
```

```
Query: 1  MNPIF-ALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFFVTLV 59
        MN F L ACFGVS+AEG +M+NLFKAAASRQPEIIGQLRSL+ILG+AF+EGTFFVTL
Sbjct: 1  MNLTLGLCLFACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLILGLIAFVEGTFFVTLA 60
```

```
Query: 60  MAFILK 65
        MAF++K
Sbjct: 61  MAFVIK 66
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 33/62 (53%), Positives = 45/62 (72%)
```

```
Query: 5  ILALGFAVMGVSIGEGILVANIAKSAARQPEMFSLQTLMTGVAFIEGTFFVLFAFTFLVR 66
        I AL A GVS+ EG L+AN+ K+A+RQPE+ +L++LM GVAFIEGTFFV F+++
Sbjct: 4  IFALALACFGVSLAEGFLMANLFKAASRQPEIIGQLRSLMILGVAFIEGTFFVTLVMAFILK 65
```

-1565-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1418**

A DNA sequence (GBSx1503) was identified in *S.agalactiae* <SEQ ID 4357> which encodes the amino acid sequence <SEQ ID 4358>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.2562(Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1419**

A DNA sequence (GBSx1504) was identified in *S.agalactiae* <SEQ ID 4359> which encodes the amino acid sequence <SEQ ID 4360>. This protein is predicted to be bacterial glycogen synthase (glgA). Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25           bacterial cytoplasm --- Certainty=0.1574(Affirmative) < succ>  
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30   The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA19591 GB:D87026 bacterial glycogen synthase [Bacillus  
                   stearothermophilus]

Identities = 220/475 (46%), Positives = 312/475 (65%), Gaps = 1/475 (0%)

35   Query: 1   MKIMFVAEAGAPFAKTGGLGDVIGALPKSLSKGHDVAVVMPYYDMVDQKFGDQIENLMY 60  
               MK++F +E APFAK+GGL DV GALPK L + G D V++P Y+ + ++ +++ +  
   Sbjct: 1   MKVLFVAVSECAPFAKSGGLADVAGALPKELRRLGIDARVMLPKYETIAPWKKMKKVAE 60

40   Query: 61   FYTDVGWRHQYVGVRKLSQDNVTFYFIDNQYYFYRGHVYGDWDDGERFAYFQLAALELME 120  
               VGWR QY GV+ L D V +YFIDN+YYF R +YG +DDGERFAYF A LE++  
   Sbjct: 61   LIVPVGWRRQYCGVEELRHDGVIIYFIDNEYFYKRPQLYGHYDDGERFAYFCRAVLEVL 120

45   Query: 121 KIDFIPDVLHVHDYHTAMIPFLLEKEYHWIQAYNNIRAVFTIHNIEFQGGQFGEMLGDLF 180  
               +I F PDV+H HD+HT M+PFLLE+Y       Y ++R VFTIHN++FQG F +L DL  
   Sbjct: 121 EIQFQPDVIHCHDWHITGMVFPFLLEQYRHELFYVDMRTVFTIHNLFQGLFPRGILEDLL 180

50   Query: 181 GVGAERYEDGTLRWNNCLNWMKAAILYSDRVTTVSPSYANEIKTPEFGKGLDQIMRMEAG 240  
               +   +   L + C+++MK A++ SD +TTVSP+Y EI+T +G+ LD ++R  
   Sbjct: 181 NLDGRYFTVDHLEFYGCVSFMKGALVASDLITTVSPTYKEETQTAYYGERLDGLLRARRD 240

Query: 241 KLSGIVNGIDSDLNPETDAFLPYHFSKSNLEGGKIKNKLALQENLGLPQDKNVPLIGIVS 300  
               L GI+NGID + NPE D FL +S   E K NK ALQ GLP+ +VPLI +V+  
   Sbjct: 241 DLLGILNGIDDEFYNPEADPFLTATYSVHTREKQLNKRALQRFGLPEWDDVPLIAMVT 300

-1566-

Query: 301 RLTDQKGFDIASELDNMLQQDIQMVILGTGYHHFEETFSYFASRYPEKLSANITFDLRL 360  
 R+T QKG D++ M+ +D+Q+V+LGTG FE+ FS A+ YP K+ I F L  
 Sbjct: 301 RMTAQKGLDLVTCVFHEMSEDMQLVVLGTGDWRFEQFFSQMAAAYPGKVGVIYIGFHEPL 360

5 Query: 361 AQQIYAASDIFMMPSAFEPCGLSQMMAMRYGSLPLVHEVGGLKDTVVAFNQFDGSGTGFS 420  
 A QIYA +D+F++PS FEPCGLSQM+A+RYG++P+V E GGL DTV ++N+ G GFS  
 Sbjct: 361 AHQIYAGADLFLIPSLFEPCGLSQMIALRYGTIPIVRETGGLNDTVQSYNEITKEGNGFS 420

10 Query: 421 FNHFSGYWLMQTLKLALEVYNDYPEAWKKLQWQAMSKDFSWDTACVAYEQLYQQL 475  
 F +F+ + ++ T++ AL Y P W++L +AM D+SW + Y+Q Y+QL  
 Sbjct: 421 FTNFAHDMLYTIRRALSFYRQ-PSVWEQLTERAMRGDYSWRRSANQYKQAYEQL 474

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1420

A DNA sequence (GBSx1505) was identified in *S.galactiae* <SEQ ID 4361> which encodes the amino acid sequence <SEQ ID 4362>. This protein is predicted to be a subunit of ADP-glucose pyrophosphorylase. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3492(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA19590 GB:D87026 subunit of ADP-glucose pyrophosphorylase  
 [Bacillus stearothermophilus]  
 Identities = 59/178 (33%), Positives = 111/178 (62%), Gaps = 1/178 (0%)

30 Query: 37 SAEIYVIDTPWLIKEMEEAQNNEPRKLRFLRLDLIVESNALAFEYTGYSNIISSIKSYY 96  
 S E+Y+++T L++ + + +N+ + ++RD + +EY+GY + I S++ Y+  
 35 Sbjct: 157 SLEMYLLETSLLLDLIADY-KNHGYYSIVDVIRDYHRSLSICEYEYSGYAAVIDSVEQYF 215

Query: 97 DANMDMLTPNKFYSLFFSNQKVYTKVKNEEATYFDKQSNVSNSQLASGSIKGYLDHSIV 156  
 ++M++L + + LF + +YTKVK+E T + ++ NV S +A+G +I+G +++S++  
 40 Sbjct: 216 RSSMELDRDVWEQLFLPSHPITVKVKDEPPTYGREGNVKRSMIANGCVIEGTVENSVL 275

Query: 157 SRNCLLEKGTTRVNSIIFPKVKIGEGATIENTIIDKCVKVASGVTLKGS�DKPLVIPK 214  
 R+ + KG V NSII K +IG+G ++ IIDK KV GV LKG+ ++P ++ K  
 Sbjct: 276 FRSVKIGKGAVVRNSIIMQKCQIGDGCVLGDGVIIDKDAKVEPGVVLKGTKEQPPIVRK 333

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1421

A DNA sequence (GBSx1506) was identified in *S.galactiae* <SEQ ID 4363> which encodes the amino acid sequence <SEQ ID 4364>. This protein is predicted to be subunit of ADP-glucose pyrophosphorylase (glgC-1). Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq

-1567-

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9765> which encodes amino acid sequence <SEQ ID 9766> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAA19589 GB:D87026 subunit of ADP-glucose pyrophosphorylase  
 [Bacillus stearothermophilus]  
 Identities = 195/352 (55%), Positives = 259/352 (73%)

15 Query: 7 MKNEMLALILAGGQGTRLGKLTQSIAPAVQFGGRYRIIDFALSNCANSINNNGVITQY 66  
 MK + +A++LAGGQG+RL LT +IAKPAV FGG+YRIIDF LSNC NSGI+ VGV+TQY  
 Sbjct: 1 MKKKCIAMLLAGGQGSRLRSLTTNIAKPAVPFGGKYRIIDFTLSNCTNSGIDTVGVLTQY 60

20 Query: 67 QPLELNTHIGNGSSWGLDGDIDSGVTVLQPYSGTEGSRWFQGTSHAIYQNIDYIDRINPEY 126  
 QPL L+++IG GS+W LD + GVTVL PYS + G +W++GT++A+YQNI+YI++ NP+Y  
 Sbjct: 61 QPLLLHSYIGISAWDLDRRNGGVTVLPPYSVSSGVKWEYGTANAVYQNIYIEQYNPDY 120

25 Query: 127 VLILSGDHIYKMNYDDMLQTHKDNLASLTVAVLDPVLPKEASRFGIMNTDSNDRIVEFEK 186  
 VL+LSGDHIYKM+Y ML H A +T++V++VP +EASRFGIMNT+ IVEF EK  
 Sbjct: 121 VLVLSGDHIYKMDYQHMLDYHIAKQADVTTISVIEVPWEEASRFGIMNTNEEMEIVEFAEK 180

30 Query: 187 PEHPKSTKASMGIIYIFDWKRLRTLIDGKNGIDMSDFGKNVIPAYLESGERVYTYNFDG 246  
 P PKS ASMGIIYIF+W L+ L N DFGK+VIP L +R + Y F+G  
 Sbjct: 181 PAEPKSNLASMGIYIFNWPLLKQYLQIDNANPHSSHDFGKDVIPMLLREKKRPFAYPFEG 240

35 Query: 247 YWKDVGTIESLWEANMEYIGEDNKLHSRDRSWKIYSKNLIAPPNFMTEANVKDSLVDG 306  
 YWKDVGT++SLWEANM+ + E+N+L DRSW+IYS N PP +++ +A V DSLV +G  
 Sbjct: 241 YWKDVGTVKSLWEANMDLLDENNELDLFDRSWRIYSVNPQPPQYISPEAEVSDSLVNEG 300

Query: 307 CFVAGNVEHSILSTNVQVKPNAIIKDSFVMSGATIGEGAKINRAIIGEDAVI 358  
 C V G VE S+L V++ A++K+S +M GA + EGA + RAI+ D++I  
 Sbjct: 301 CVVEGTVERSFLFQGVRIKKGAVVKESVIMPGAASVSEGAYVERAIVTPDSII 352

There is also homology to SEQ ID 2660.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1422**

A DNA sequence (GBSx1507) was identified in *S.agalactiae* <SEQ ID 4365> which encodes the amino acid sequence <SEQ ID 4366>. Analysis of this protein sequence reveals the following:

45 Possible site: 37  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2844 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAA78440 GB:Z14057 1,4-alpha-glucan branching enzyme [Bacillus  
 caldolyticus]  
 Identities = 272/616 (44%), Positives = 371/616 (60%), Gaps = 14/616 (2%)

Query: 6 ELYTFGIGENFHLQNYLGVHSENGSFC----FRVWAPNAENVQVIGDFTDWRNRPLQMNK 61  
 E+Y F G + G H G F VWAP+A V+++G F DW + K

-1568-

Sbjct: 10 EVYLFHEGRLYQSYELFGAHVIRGGGAVGTRFCVWAPHAREVRLVGSFNDWNGTNSPLTK 69  
 Query: 62 -NQAGVWEANSLDAREGDLYKYLVTCKGGQVVEKIDPMVYMERRPGTASVIKVLNRKKN 120  
 N GVV + EG LYKY + G+V+ K DP A Y E RP TAS++ L+ +W  
 Sbjct: 70 VNDEGVWTVVVPENLEGHLYKYEIITPDGRVLLKADPYAFYSELRPHTASIVYDLKGYEW 129  
 Query: 121 EDGLWMGRKRRLGFGQKRPINITYEVHAGSWKKDDFGHPMTFSQKDYLIPIYLVEMNHYTHVE 180  
 D W +++R +P+ IYE+H GSWKK G T+ ++ D LIPY++E +TH+  
 Sbjct: 130 NDSPWQRKKRRKRIYDQPMVIYELHFGSWKKKPDGRFITYREMADELIPYVLERGFTHIE 189  
 Query: 181 FMPLMAHPLDMSWGYQLMGYFAFEHTYGTPEEFQDFVEACHKNNIGVLVDWVPGHFQND 240  
 +PL+ HPLD SWGYQ GY++ YGTP +F FV+ CH+ +GV++DWVPGHF ++  
 Sbjct: 190 LLPLVEHPLDRSWGYQGTGYYSVTSRYGTPHDFMYFVDRCHQAGLGVIIDWVPGHFCKDA 249  
 Query: 241 DALAYFDGTATYEQNHDRAHNYRWGALNFDLGKNQVQSFLISSALFWIEHYHIDGIRVD 300  
 L FDG TYEY N NY WG NFDLGK +V+SFLIS+ALFW+E+YH+DG RVD  
 Sbjct: 250 HGLYMFDGAPTYEYANEKORENYVWGTANFDLGKPEVRSFLISNALFWLEYVHVDGFRVD 309  
 Query: 301 AVSNMLYLDYDEGPWEANQFGDNRNLEGYHFLRKINKVIKERHPNVMMIAEESTASTPIT 360  
 AV+NMLY ++ +E N FLR+LN+ + PNV MIAE+ST +T  
 Sbjct: 310 AVANMLYWPNDRLYE-----NPHYAVEFLRQLNEAVFAYDPNVMMIAEDSTDWPRVT 361  
 Query: 361 KDLESGLGFDKWNMGWMNDILRFYEDPLRYQYDFNLVTFSEFMYIPNENFVLAFSHDE 420  
 GGLGF++KWNMGWMND+L++ E P R+Y N V+FS +Y ++ENF+L FSHDE  
 Sbjct: 362 APTYDGGGLGFNYKWNMGWMNDMLKYMETPPHERKYAHNQVSFSLLYAYSENFILPFSHDE 421  
 Query: 421 VVHGKKSMMHKMWGDRYNQFAGLRNLYAYQMCHPGKLLFMGSEFGQFLEWKYNDQLEWE 480  
 VVHGKKS+++KM G +FA LR LY Y M HPGKLLFMGSEF QF EWK+ ++L+W  
 Sbjct: 422 VVHGKKSLLNKMPPGSYEEKFAQLRLLYGYMMAHPGKLLFMGSEFAQFDEWKFAEELDWW 481  
 Query: 481 NLNDMMNQMKORYTKQLNQFYKDKHCLWRIDDSFDGLEIIDADNKSETVLSFIRKDDK-G 539  
 + ++++KM Y KQL YK +K + +D G E ID N +++ SFIR+ K G  
 Sbjct: 482 LDFELHRKMDEYVKQLIACYKRYKPFYELDHDPRGFWDVHNAEQSIFSIFIRRGKKEG 541  
 Query: 540 DLLLCVFNMTFVERPNFTIGVPQAGIYEEVLNTEMEEFGGVWKNHNPVTKTQVATWKDYD 599  
 D+L+ V N T ++ + VP Y EVLN++ EFGG + +  
 Sbjct: 542 DVLVIVCNFTNQAYDDYKVSVPPLAPYREVLNSDAEFGGSGHVNGKRLPAFSEPFHGKP 601  
 Query: 600 HTLSFTLPALGASVWR 615  
 + + T+P G S+ R  
 Sbjct: 602 YHVRMTIPPFGISILR 617

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### 45 Example 1423

A DNA sequence (GBSx1508) was identified in *S.agalactiae* <SEQ ID 4367> which encodes the amino acid sequence <SEQ ID 4368>. This protein is predicted to be pullulanase (pula). Analysis of this protein sequence reveals the following:

Possible site: 45  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3194(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44685 GB:U67061 pullulanase [Bacteroides thetaiotaomicron]  
 Identities = 223/597 (37%), Positives = 331/597 (55%), Gaps = 55/597 (9%)  
 Query: 139 EYSETKTAFRLWAPTAERVELILYHSTDETASVSKVLSMKRGTA VNYKNHKENTHGVWFT 198

-1569-

EY+ T F LW+PTA+ V L+LY + E + + M+ G G W  
 Sbjct: 46 EYTPEATKFTLWSPTADEVRLMLYEA-GEKGHAYETVKMQSGE-----EGTWTA 93  
 Query: 199 ELEGNYNYQAYTYRVVYRRRTFKITRDPYSIATTANGKRSIVIAPEALTPEGFKLSHGKE 258  
 + + + Y+ V + T + A NGKR+ +I ++ P+G++ +  
 Sbjct: 94 VVSKDLIGKFYTFNVKIDDKWQGDTPGINARAVGVNGKRAALIDWQSTNPDGWE----SD 149  
 Query: 259 AKWRLENPNQAVIYEMHVRDFSISSETSGVKTDYHGKFKGLHQKGTVNQHGDKTTFDYVQD 318  
 + L++P +IYEMH RDFS+ TSGVK GK+ L + GT+N T D++ +  
 Sbjct: 150 TRPPLKSPADMIYEMHHRDFSVDSTSGVKNK--GKYLALTEHGTMSDKLLTGIDHLIE 207  
 Query: 319 LGVNYIQLQPIFDHHQTFDDDD-GHYAYNWGYDPENYNVPEASFSSNPHEPATRILELKSA 377  
 LGV ++ L P FD+ + +YNWGYDP+NYNVP+ S++++P++PATR+ E K  
 Sbjct: 208 LGVTHVHLLPSFDYASVDETRLNENSYNWGYDPQNYNVPDGSYATDPYQPATRVKEFKQM 267  
 Query: 378 IQAYHDAGIGVIMDVVYNHTFSSTDSAFQLTVPDYRRMNHNGTFQNGSGCGNETASEKE 437  
 +QA H AGI VIMDVVYNHTF++ +S F+ TVP Y+YR + T NGSGCGNETASE+  
 Sbjct: 268 VQALHKAGIRVIMDVVYNHTFNTDESINFERTVPGYFYRQKEDKTLANGSGCGNETASERL 327  
 Query: 438 MCRKYILDSVLVWVKEYNIDGFRFDLMGLHDVETMNIIRNELNKIDPRILVYEGGWDMA 497  
 M RK++++SVLYW+KEY++DGFRFDLMG+HD+ETMN IR +N +DP I +YEGGW A  
 Sbjct: 328 MMRKFMVESVLYWIKYHVDGFRFDLMGIHDIETMNEIRKAVNAVDPTICIIYEGGWAAEA 387  
 Query: 498 GLTPQNK-AKKDNAYQMPGIGFFNDVVRDAV---KGAEIYGEFKKGLVSGNSTEDIVAKG 553  
 P + A K N Q+PG+ F+D++RD + G + G F G+ G E V G  
 Sbjct: 388 PQYPADSLAMKGNIAQIPGVAVFSDELRLDGLCGPVGDKRKGAFLAGIPGG---EMSVKFG 444  
 Query: 554 ILGSDE-----LVSYI-----DPSQVLNYVEAHDNYNLNDLLWELHPNDNEKQHIYR 600  
 I G+ E V+Y P Q+++YV HD L D L P+ +Q I  
 Sbjct: 445 IAGAIEHPQVQCDSVNYTQKPWAKQVQMISYVSCHDGLCLVDRLKASMPDITPEQLIRL 504  
 Query: 601 VEVASAMNLLMQGMAFMQLGQEFRLTKCYPTGDKGQLTQADKERAMNSYNAPDQVNVQNW 660  
 ++A A+ QG+ F+ G+E +R DK+ NSY +PD VN ++W  
 Sbjct: 505 DKLAQAVVFTSQGIPFIYAGEEIMR-----DKQGVNDSYKSPDAVNAIDW 549  
 Query: 661 DNVTFHKSTINFIRKIITLKTNSPYFSYSSFEIRKHFVESAQYHSGFISFTVEEH 717  
 T + +++I L+ + P F ++RKH+ + S I+F +++H  
 Sbjct: 550 RRKTTADVFMYYKRLIDLRKSHPAFRMGDAGQVRKHLEFLPVE-GSNLIAFRLKDH 605

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1424

A DNA sequence (GBSx1509) was identified in *S.agalactiae* <SEQ ID 4369> which encodes the amino acid sequence <SEQ ID 4370>. Analysis of this protein sequence reveals the following:

Possible site: 40  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2368(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12492 GB:Z99107 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 151/293 (51%), Positives = 193/293 (65%), Gaps = 5/293 (1%)  
 Query: 5 KKARLIYNPTSGQEIMKKNVAEVLIDILEGFGYETSAFQTTPTKNSARDEATRAAQAGFDL 64  
 K+AR+IYNPTSG+EI KK++A+VL E GYETS TT A A AA FDL  
 Sbjct: 2 KRARIYNPTSGREIFKKHLAQVLQKFEQAGYETSTHATT-CAGDATHAAKEAALREFDL 60  
 Query: 65 IVAAGGDGTINEVVNGIAPLKRPKMAIIPGTITNDFARALKIPRGNPIEATKLIGKNQI 124  
 I+AAGGDGTINEVVNG+APL RP + +IP GTTNDFARAL IPR + ++A +

-1570-

Sbjct: 61 IIAAGDGTINEVVNGLAPLDNRPTLGVIPVGTITNDFARALGIPREDILKAADTVINGVA 120

Query: 125 VKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRKVPVK 184  
+DIGQ YFINIA G LTELTY VPS+LKT G LAY KG+E+LP +R V+

5 Sbjct: 121 RPIDIGQVN-GQYFINIAGGRLTELTYDVPSKLKTMGLQAYYLKGMEMLPRLPTEVE 179

Query: 185 ITHDKGEFIGDASMI FVAITNSVGGFEQIAPDAKLDDGKFTLILVKTANLIEIMHLIRLV 244  
I +D F G+ + V +TNSVGGFE++APD+ L+DG F L+++K ANL E + + +

10 Sbjct: 180 IEYDGKLFQGEIMLFLVTLTNSVGGFEKLAPDSSLNDGMFDLMILKKANLAEFIRVATMA 239

Query: 245 LAGGKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHI 297  
L G+HIND+ + Y K + + + E+M +NLDGEYGG P NL HI

Sbjct: 240 LR-GEHINDQHIIYTKANRVKVN--VSEKQLNLDGEYGGMLPGEFVNLYRHI 289

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4371> which encodes the amino acid sequence <SEQ ID 4372>. Analysis of this protein sequence reveals the following:

Possible site: 40  
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2501(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 272/334 (81%), Positives = 300/334 (89%)

Query: 1 MKKQKARLIYNPTSGQEIMKKNVAEVL DILEGFGYETSAFQTTPTKNSARDEATRAAQA 60  
MKKQ +ARLIYNPTSGQE+M+K+V EVLDILEGFGYETSAFQTT KNSA +EA RAA+A

30 Sbjct: 1 MKKQLRARLIYNPTSGQELMRKSVPEVL DILEGFGYETSAFQTTAQKNSALNEARRAAKA 60

Query: 61 GFDLIVAAGDGTINEVVNGIAPLKRPPKMAI IPTGTITNDFARALKIPRGNPIEATKLIG 120  
GFDL++AAGDGTINEVVNGIAPLK+RPKMAI IPTGTITNDFARALK+PRGNP +A KLIG

35 Sbjct: 61 GFDLLIAAGDGTINEVVNGIAPLKKRPKMAI IPTGTITNDFARALKVPRGNPSQAAKLIG 120

Query: 121 KNQIVKMDIGQAQEDNYFINIAAAGSLTELTYSVPSQLKTTFGYLAYLAKGVELLPRVRK 180  
KNQ ++MDIG+A++D YFINIAAAGSLTELTYSVPSQLKT FGYLAYLAKGVELLPRV

Sbjct: 121 KNQTIQMDIGRAKIDTYFINIAAAGSLTELTYSVPSQLKTMFGYLAYLAKGVELLPRVSN 180

40 Query: 181 VPKITHDKGEFIGDASMI FVAITNSVGGFEQIAPDAKLDDGKFTLILVKTANLIEIMHL 240  
VPKITHDKG F G SMIF AITNSVGGFE IAPDAKLDDG FTLIL+KTANL EI+HL

Sbjct: 181 VPKITHDKGVFEGQVSMIFAAITNSVGGFEMIAPDAKLDDGMFTLILIKTANLFEIVHL 240

45 Query: 241 IRLVLGAGKHINDKRVEYIKTSYLTIEPLSDERMMINLDGEYGGDAPITLANLKNHIRFF 300  
+RL+L GKKHI D+RVEYIKTS + IEP +RMMINLDGEYGGDAPITL NLKNHI FF

Sbjct: 241 LRLILDGKKHITDRRVEYIKTSKIVIEPQCGKRMMINLDGEYGGDAPITLENLKNHITFF 300

Query: 301 ANTDEISDDALVLDKDELAIEATAQKFANEVDDL 334  
A+TD ISDDALVLD+DEL IE I +KFA+EV+DL

50 Sbjct: 301 ADTDLISDDALVLDQDELEIEIVKKFAHEVEDL 334

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1425

- 55 A DNA sequence (GBSx1510) was identified in *S.agalactiae* <SEQ ID 4373> which encodes the amino acid sequence <SEQ ID 4374>. This protein is predicted to be DNA ligase (ligA-1). Analysis of this protein sequence reveals the following:

Possible site: 16  
>>> Seems to have no N-terminal signal sequence

-1571-

INTEGRAL Likelihood = -0.27 Transmembrane 363 - 379 ( 363 - 379)

----- Final Results -----

5 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9763> which encodes amino acid sequence <SEQ ID 9764> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12482 GB:Z99107 similar to DNA ligase [Bacillus subtilis]  
Identities = 346/657 (52%), Positives = 462/657 (69%), Gaps = 8/657 (1%)

15 Query: 2 ENRMNELVSLNQYAKYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHR VGG 61  
+ R EL +N+Y+ EYYT D P+V D++YD+L +EL+ +E++HP+ P+SPT RVGG  
Sbjct: 7 KQRAEELRRTINKYSY EYYTLD E P S V P D A E Y D R L M Q E L I A I E E H P D L R T P D S P T Q R V G G 66

20 Query: 62 LVLEGFEKYQHEYPYLSLQDAFSKEELIAFDKRVKAEF-PTAAYMAELKIDGLSVSLTYV 120  
VLE F+K H P+ SL +AF+ ++L FD+RV+ AY ELKIDGL+VSL Y  
Sbjct: 67 AVLEAFQKVTHGTPMLS LGNAFNADDLRDFDRVRVQSVGDDVAYNVELKIDGLAVSLRYE 126

25 Query: 121 NGVLQVGATRGDGNIGENITENLKRVDIPLHLQSLDITVRGECYLPKESFEAINIEKR 180  
+G GATRGDG GE+ITENLK + +IPL +++ L I VRGE Y+PK SFEA+N E+  
Sbjct: 127 DGYFVRGATRGDGTGEDITENLKTIRNIPLKMNRELSIEVRGEAYMPKRSFEALNEERI 186

30 Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQK--ETQDDVLKELESYG 238  
N E+ FANPRNAAAG+LRQL+ I AKR L F+Y A + ETQ L L+ G  
Sbjct: 187 KNEEPPFANPRNAAAGSLRQLDPKIAAKRNLDIFVYSIAELDEMGEVETQSQGLDFLDELG 246

35 Query: 239 FSVNHHRLISSMEKIWDIFIQTIEKDRVSLFPYDIDGIVIKVNSIAMQEELGFTVKAPRWA 298  
F N R S+E++ I ++ R LPY+IDGIVIKV+S+ QEELGFT K+PRWA  
Sbjct: 247 FKTNQERKKCGSIEEVITLIDELQAKRADLPYEIDGIVIKVDSLDDQEELGFTAKSPRWA 306

40 Query: 299 IAYKFPAEEKEAEILSVDWTVGRTGVVPTANLTPVQLAGTTVSRATLHNVDIYAEKDIR 358  
IAYKFPAEE ++L ++ VGRTGV+TPTA L PV++AGTTVSRA+LHN D I EKDIR  
Sbjct: 307 IAYKFPAEEVVTKLLDIELNVGRTGVITPTAILEPVKVAGTTVSRASLHNEEDLIKEKDIR 366

45 Query: 359 IGDTVVVYKAGDIIPAVLNVVMSKRNQQEVML-IPKLCPSGSELVHFEGEVALRCINPL 417  
I D VVV KAGDIIP V+NV++ +R +E +P CP CGSELV EGEVALRCINP  
Sbjct: 367 ILDKVVVKAGDIIPEVNVVLVDQRTGEEKEFSMPTECECGSELVRIEGEVALRCINPE 426

50 Query: 418 CPNQIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEK 477  
CP QI+E L HF SR+AMNI G G ++ +LF+ +L+ +VAD+Y+L+ E ++ L+ + EK  
Sbjct: 427 CPAQIREGLIHFVSRNAMNIDGLGERVITQLFEENLVRNVADLYKLTKEKRVIQLERMGEK 486

55 Query: 478 SATKIYHAIQSSKENS AEKLLFGLGIRHVGSKASRLLEEFGNLRQLSQASQESIASIDG 537  
S + +IQ SKENS E+LLFGLGIR +GSKA++ L F +L L +AS+E + ++D  
Sbjct: 487 STENLISSIQSKENSLEKLLFGLGIRFIGSKAAKTLAMHFESLENLKKASKEELLAVDE 546

Query: 538 LGGVIAKSLHTFFEKEEVDKLLLEELTSYNVNFNYLG----KRVSTDAQLSGLTVVLTGKL 593  
+G +A ++ T+F KEE+ +LL EL VN Y G K +D+ +G T+VLTGKL  
Sbjct: 547 IGEKMAVITYTFHKEEMLELLNELQELGVNTLYKGPVKVKAEDSDSYFAGKTIVLTGKL 606

Query: 594 EKMRNEAKEKLN LGAKVTGVS VKTDLIVAGSDAGSKLTKAQDLGITIQDEDWLL 650  
E+++RNEAK +++ LG K+TGSVSK TDL++AG AGSKLTKAQ+L I + +E+ L+  
Sbjct: 607 EELSRNEAKAQIEALGGKLTGVS VKTDLVIAGEAAGSKLTKAQELNIEVWNEEQLM 663

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4375> which encodes the amino acid sequence <SEQ ID 4376>. Analysis of this protein sequence reveals the following:

60 Possible site: 61  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -0.43 Transmembrane 363 - 379 ( 363 - 379)

-1572-

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 472/652 (72%), Positives = 556/652 (84%)

10 Query: 1 MENRMNELVSLNQAKEYYTQDNPTVSDSQYDQLYRELVELEKQHPENILPNSPTHRVG 60  
 M+ R+ EL LLN+Y +YYT+D P+VSDS YD+LYRELV LE+ +PE +L +SPT +VG  
 Sbjct: 1 MKKRIKELTDLNRYRYDYTKDAPSVSDSDYDKLYRELVTLEQSYPEYVLQDSPTQQVG 60

15 Query: 61 GLVLEGFEKYQHEYPYLSLQDAFSKEELIAFDKRVKAEFPTAAYMAELKIDGLSÚSLTYV 120  
 G +L+GFEKY+H+YPL+SLQDAFS+EEL AFDKRVKAEFP A Y+AELKIDGLS+SL+Y  
 Sbjct: 61 GTILKGFEKYRHQYPLFSLQDAFSREELDAFDKRVKAEFPNATYLAELKIDGLSISLSYE 120

20 Query: 121 NGVLQVGATRGDGNIGENITENLKRVDIPLHLQSLDITVRGECYLPKESFEAINIEKR 180  
 NG LQVGATRGDGNIGENITEN+K++ DIP L + L ITVRGE Y+ ++SF+AIN ++  
 Sbjct: 121 NGFLQVGATRGDGNIGENITENIKIKIDIPYQLSEPLTITVRGEAYMSRQSFKAINEARQ 180

25 Query: 181 ANGEQEFANPRNAAAGTLRQLNTGIVAKRKLATFLYQEASPTQKETQDDVLKELESYGFS 240  
 NGE EFANPRNAAAGTLRQL+T +VAKR+LATFLYQEASPT + Q++VL EL GFS  
 Sbjct: 181 ENGETEFANPRNAAAGTLRQLDTSVAKRQLATFLYQEASPTARNQNEVLAEADLGF 240

30 Query: 241 VNHRLISSSMEKIWDFIQTIEKDRVSLPYDIDGIVIKVNSIAMQEELGFTVKAPRWAI 300  
 VN + ++SSM++IWDFI+TIE R L YDIDG+VIKVN+AMQEELGFTVKAPRWAI  
 Sbjct: 241 VNPYYQLTSSMDEIWDFIKTIEAKRQLAYDIDGVVIKVN+SLAMQEELGFTVKAPRWAI 300

35 Query: 301 YKFPAAEKEAEILSVDWTVGRGVVPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360  
 YKFPAAEKEAEILSVDWTVGRGVVPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG  
 Sbjct: 301 YKFPAAEKEAEILSVDWTVGRGVVPTANLTPVQLAGTTVSRATLHNVDYIAEKDIRIG 360

40 Query: 361 DTVVVYKAGDIIPAVLNVVMSKRNOQEVMLIPKLCPCGSELVHFEFEVALRCINPLCPN 420  
 DTV+VYKAGDIIPAVLNVVMSKRNOQEVMLIPKLCPCGSELVHFE EVALRCINPLCP+  
 Sbjct: 361 DTVIVYKAGDIIPAVLNVVMSKRNOQEVMLIPKLCPCGSELVHFEFEVALRCINPLCP 420

45 Query: 421 QIKERLAHFASRDAMNITGFGPSLVEKLFDAHLIADVADIYRLSIENLLTLDGIKEKSAT 480  
 I+ L HFASRDAMNITG GP++VEKLF A + DVADIY+L+ E+ + LDGIKEKSA  
 Sbjct: 421 LIQRSLEHFASRDAMNITGLGPAIVEKLFFLAGFVHDVADIYQLTKEDFMQLDGIKEKSAD 480

50 Query: 481 KIYHAIQSSKENSAAEKLFLGIRHVGSKASRLILLEEFGNLRQLSQASQESIASIDGLGG 540  
 K+ AI++SK NSAEKLFLGIRH+GSK SRL+LE +G++ L A +E IA IDGLG  
 Sbjct: 481 KLLAAIEASKNSAEKLFLGIRHIGSKVSRILEVYGDISALLTAKEEETARIDGLGS 540

55 Query: 541 VIAKSLHTFFEEKEEVDKLEELTSYNVNFNYLGKRVSTDAQLSGLTVVLTGKLEKMTRNE 600  
 IA+SL +FE++ L++EL + VN +Y G++V++DA L GLTVVLTGKL ++ RNE  
 Sbjct: 541 TIAQSLTQYFEQKTAAILVDELKTAGVNMHYSQKVNSDAALFGLTVVLTGKLNQLNRNE 600

60 Query: 601 AKEKLQNLGAKVTGVSVKKTDLIVAGSDAGSKLTKAQDLGITIQDEDWLLNL 652  
 AK+KL+ LGAKVTGVSVKKTDL++AGSDAGSKL KA+ LGI I+DEDWL L  
 Sbjct: 601 AKDKLEALGAKVTGVSVKKTDLVIAGSDAGSKLEKAKSLGIRIEDEDWLRQL 652

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 55 Example 1426

A DNA sequence (GBSx1511) was identified in *S.agalactiae* <SEQ ID 4377> which encodes the amino acid sequence <SEQ ID 4378>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

60

INTEGRAL	Likelihood = -5.63	Transmembrane	110 - 126 ( 108 - 128)
INTEGRAL	Likelihood = -2.13	Transmembrane	142 - 158 ( 141 - 159)
INTEGRAL	Likelihood = -1.12	Transmembrane	75 - 91 ( 75 - 93)

-1573-

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.3251(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus  
 plantarum]  
 Identities = 56/158 (35%), Positives = 91/158 (57%), Gaps = 8/158 (5%)  
 Query: 13 AIVTAIYIVLTITPPFNALAYGAYQFRVSEMLNFLAFYHRKYLFAVTLGCMISNLYSFG- 71  
 A+V A+Y+VL + P ++A GA QFRVSE LN LA ++RKY++ + G ++ + + G  
 15 Sbjct: 13 ALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRYKIWGIVAGVILFDFAGPGA 72  
 Query: 72 -MIDVFVGGGSTLLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAASMITVAVELKI 130  
 +++V GGG +LL + + T L + K L+N A F S F A MIT+ +  
 Sbjct: 73 SLLNVLFGGQSLLALLVLTWLAPKL-KTVWQRMILLNIALFTVSMFMIALMITM-----M 126  
 20 Query: 131 VAGLPLLLTTLTAVGELASLLVGAVLVDKLSRHVDFT 168  
 +G+ T+LTTA+ EL + + A ++ L R + F+  
 Sbjct: 127 SSGVAFWPTYLTALSELIIMSITAPIMYSILDRVLFHS 164

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4379> which encodes the amino acid  
 sequence <SEQ ID 4380>. Analysis of this protein sequence reveals the following:

Possible site: 32  
 >>> Seems to have an uncleavable N-term signal seq  
 30 INTEGRAL Likelihood = -4.41 Transmembrane 75 - 91 ( 70 - 94)  
 INTEGRAL Likelihood = -3.82 Transmembrane 12 - 28 ( 8 - 28)  
 INTEGRAL Likelihood = -2.28 Transmembrane 141 - 157 ( 140 - 158)  
 INTEGRAL Likelihood = -0.64 Transmembrane 110 - 126 ( 110 - 126)  
 INTEGRAL Likelihood = -0.59 Transmembrane 55 - 71 ( 54 - 73)

## 35 ----- Final Results -----

bacterial membrane --- Certainty=0.2763(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 114/167 (68%), Positives = 137/167 (81%), Gaps = 1/167 (0%)  
 Query: 1 MNTFTTRODYAHMAIVTAIYIVLTITPPFNALAYGAYQFRVSEMLNFLAFYHRKYLFAVTL 60  
 M T DY H+ +V A+Y+VLITIPP NAI+YG YQFR+SEM+NFLAFYHRKY+ AVTL  
 45 Sbjct: 1 MTKLTVHDYVHIGLVAALYVVLITITPPLNAISYGMYPRISEMMNFLAFYHRKYIIAVTL 60  
 Query: 61 GCMISNLYSFGMIDVFVGGGSTLLFVYLGTILFKQYQKDYLFNGLINKAFFFFSFFFAAS 120  
 GCMI+N YSFG+IDVFVGGGSTL+FV LG ILF +YQKDYLFNG+ NKAF +FSFFFA S  
 50 Sbjct: 61 GCMIANFYSFGLIDVFVGGGSTLIFVTLGVILFSKYQKDYLFNGIFNKAFVYFSFFFAATS 120  
 Query: 121 MITVAVELKIVAGLPLLLTTLTAVGELASLLVGAVLVDKLSRHVD 167  
 M VA+EL G P LLTW TTA+GEL SLL+G++++DKLS+ + F  
 Sbjct: 121 MFNVAIELYFF-GAPFLLTWFTTALGELVSLIGSLIIDKLSQRISF 166

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

**Example 1427**

A DNA sequence (GBSx1513) was identified in *S.agalactiae* <SEQ ID 4381> which encodes the amino  
 acid sequence <SEQ ID 4382>. Analysis of this protein sequence reveals the following:

-1574-

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -11.20	Transmembrane	255 - 271 ( 245 - 281)
INTEGRAL	Likelihood = -10.72	Transmembrane	141 - 157 ( 132 - 165)
INTEGRAL	Likelihood = -8.17	Transmembrane	189 - 205 ( 185 - 208)
INTEGRAL	Likelihood = -7.01	Transmembrane	36 - 52 ( 33 - 60)

----- Final Results -----

bacterial membrane	---	Certainty=0.5479(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]

Identities = 83/276 (30%), Positives = 157/276 (56%), Gaps = 3/276 (1%)

Query: 17 RPIQVFMRFQSAEMDLASIAVAYYLLVTAFFPLLVAANIFPYFHINVSDDLMLQKNLP 76

R I+ H +AE+ S++ VAYYLL++ FPLL+ N+ PY I+ + +L + + +P

Sbjct: 15 RFIETQSHMVTAEIGNSSVVVAYYLLSLFPLLIAVGNVLPYLRIDPNSVLPYIAEAIP 74

Query: 77 KNIYEPASRLAVDAFSKPSTGILGFASLTAFWTMSKSLTSLQKAINKAYGVDQHRDFVIS 136

K++Y+ ++ S G+L ++L AFW+ S+S+ +LQ A+NKA+GV+Q ++F++

Sbjct: 75 KDVKYKNEPAIRSLLTQRSGLLSVSALAAFWSASQSINALQAMNKAFGVEQRKNFILV 134

Query: 137 RLVGVTGLIILFLLTFVLIFSTFSPVQLIIVNMYDLGDTLTAWLLNLAQPVTFITIFL 196

R+V L+ + + V++ + +++++ ++ ++ L P+T + + +

Sbjct: 135 RVVSFLVILLFMVAIVGVVILGLGQYIIEQLQPIFYHSTVIDTFQALKWPLTTVVLLV 194

Query: 197 GIGILYFILPNARIRKVRYPVPGTLFSTFVIGFFSNLISQVVLNRVEKMVDIKTFGSVVI 256

+ ++Y ++PN ++ +R ++PG +FST S + YV ++ + GS +

Sbjct: 195 IMCLIVAVVPNRKL-SLRSLPGAIFSTVGVWMLLSQIFGLYVKYFSSRIASYQIIGSFI- 252

Query: 257 FILMLWFIFLAHIMILGAILNASVQEIATGKIESRR 292

ILMLW F A I+ILGAI+NA V E G E ++

Sbjct: 253 -ILMLWLNFAATIIILGAIVNAVVDLYLXGXKEKKQ 287

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4383> which encodes the amino acid sequence <SEQ ID 4384>. Analysis of this protein sequence reveals the following:

Possible site: 53

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -12.58	Transmembrane	141 - 157 ( 132 - 168)
INTEGRAL	Likelihood = -12.15	Transmembrane	189 - 205 ( 177 - 210)
INTEGRAL	Likelihood = -11.68	Transmembrane	256 - 272 ( 245 - 280)
INTEGRAL	Likelihood = -7.54	Transmembrane	36 - 52 ( 33 - 60)

----- Final Results -----

bacterial membrane	---	Certainty=0.6031(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the databases:

&gt;GP:CAA68244 GB:X99978 citrulline cluster-linked gene [Lactobacillus plantarum]

Identities = 53/170 (31%), Positives = 92/170 (53%), Gaps = 11/170 (6%)

Query: 1 MTKLTVHDYVHIGLVAALYVVLITPPLNAISYGMVQFRISEMMNFLAFYHRKYIIAVTL 60

MT+ + ++ LVAA+YVVL + P +++ G QFR+SE +N LA ++RKYI +

Sbjct: 1 MTQSKIRPWIINALVAAMYVVLCLGPAAFSLASGAIQFRVSEGLNHLAVFNRYIWIWIVA 60

Query: 61 GCMIANFYSGF--LIDVFVGGGSTLIFVTLGVLFSKYQKDYLFNGIFNKAFFVYFSFFFA 118

G ++ + + G L++V GGG +L+ + + L K + ++ + + + F

Sbjct: 61 GVILFDAGPGASLLNVLFGGGQSLALVLTLWAPKLKT-----VWQRMMLNIA-LFT 113

Query: 119 TSMFNVA--IELYFFGAPFLLTWFTTALGELVSLIGSLIIDKLSQRISF 166

-1575-

SMF +A I + G F T+ TTAL EL+ + I + I+ L + + F  
 Sbjct: 114 VSMFMIALMITMSSGVAFWPTYLTALSELIIMSITAPIMYSLDRVLHF 163  
 lGB:AF071085 Orfde2 [Enterococcus faecalis] 176 2e-43

>GP:AAC35915 GB:AF071085 Orfde2 [Enterococcus faecalis]  
 Identities = 90/271 (33%), Positives = 155/271 (56%), Gaps = 3/271 (1%)

Query: 19 IQVFMRLQSAEMDLSAIAVAYYLILTAFFPLIVIAANIFFPYLNIDIADLLRLMKQNLPKD 78  
 I+ H+ +AE+ S++ VAYYL+L+ FPL++ N+ PYL ID +L + + +PKD  
 Sbjct: 17 IETTQSHMVTAEIGNSSVVVAYYLLLSLFLLIAGVGNVLPYLRIDPNSVLPYIAEAIPKD 76

Query: 79 IFRPASAIVENIFSKPSGVLGVATLTGLWTMSRSLTSLQKAINKAYGASQHRDFFIGHL 138  
 +++ + ++ ++ SG +L V+ L W+ S+S+ +LQ A+NKA+G Q ++F + +  
 Sbjct: 77 VYKNLEPAIRSLLTQSGGLLSVSALAAFWSASQSINALQNAMNKAFGVEQRKNFILVRV 136

Query: 139 VGLLTSLIILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITTIFLLLIQIPITVLIIFVGL 198  
 V L L+ + + ++ + I++L +H S ++ F L P+T +++ V +  
 Sbjct: 137 VSFLVILLFMVAIVGVVILGLQYIIELLQPIFYHSTSVIDTFQALKWPLTTVVLVIM 196

Query: 199 MLLYFLLPNVKIKIRYILPGTLFTSFVMTFSLNLVGNVYVYVVERMVDIKMFGSVMIFI 258  
 L+Y ++PN K+ +R ILPG +F++ LS + G YV Y R+ ++ GS I  
 Sbjct: 197 CLIYAVVPNRKL-SLRSILPGAIFSTVGWMLLSQIFGLYVKYFSSRIASYQIIGS--FII 253

Query: 259 IMLWFIFLARILILGAIFNATYQEMSLGKLE 289  
 +MLW F A I+ILGAI NA E G E  
 Sbjct: 254 LMLWLNFAATIIILGAIVNAVVD EYLXGXKE 284

An alignment of the GAS and GBS proteins is shown below.

Identities = 188/302 (62%), Positives = 244/302 (80%)

Query: 1 MKLKKFFEDLLAKLEYRPIQVFMRFQSAEMDLSAIAVAYYLLVTAFFPLLVIAANIFFPYF 60  
 M KK+F+ +L+K +Y PIQVFMRF QSAEMDLSAIAVAYYL++TAFPL+VIAANIFFY  
 Sbjct: 1 MAEKKWFDKVLKWKQYEPPIQVFMRLQSAEMDLSAIAVAYYLILTAFFPLIVIAANIFFPYL 60

Query: 61 HINVSDLLSLMQKNLPKNIYEPASRLAVDAFSKPSTGILGFASLTAFWTMSKSLTSLQKA 120  
 +I+++DLL LM++NLPK+I+ PAS + + FSKPS +LG A+LT WTMS+SLTSLQKA  
 Sbjct: 61 NIDIADLLRLMKQNLPKDIFRPASAIVENIFSKPSGVLGVATLTGLWTMSRSLTSLQKA 120

Query: 121 INKAYGVDQHRDFVISRLVGVGTGLIILFLLTFVLIFSTFSKPVLQIIVNMYDLGDTLTA 180  
 INKAYG QHRDF I LVG+ T LIILFLL F LIFS FSK +Q++ Y L D +T  
 Sbjct: 121 INKAYGASQHRDFFIGHLVGLLTSLIILFLLAFALIFSIFSKAAIQVLDKHYHLSDNITT 180

Query: 181 WLLNLAQPVTFLTIFLGIGILYFILPNARIRKRVYVIPGTLFSTFVIGFFSNLISQYVLN 240  
 L L QP+T L IF+G+ +LYF+LPN +I+K+RY++PGTLF++FV+ F SNL+ YV+  
 Sbjct: 181 IFLLLIQIPITVLIIFVGLMLLYFLLPNVKIKIRYILPGTLFTSFVMTFSLNLVGNVYVY 240

Query: 241 RVEKMVDIKTFGSVVIFILMLWFIFLAHIMILGAILNASVQEIATGKIESRRGDIMSLIQ 300  
 VE+MVDIK FGSV+IFI+MLWFIFLA I+ILGAI NA+ QE++ GK+E R GD+++++  
 Sbjct: 241 NVERMVDIKMFGSVMIFITIMLWFIFLARILILGAIFNATYQEMSLGKLEGRSGDMIAILK 300

Query: 301 KS 302  
 K+  
 Sbjct: 301 KT 302

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1428

A DNA sequence (GBSx1514) was identified in *S. agalactiae* <SEQ ID 4385> which encodes the amino acid sequence <SEQ ID 4386>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

-1576-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4200(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

## 10 Example 1429

A DNA sequence (GBSx1515) was identified in *S.agalactiae* <SEQ ID 4387> which encodes the amino acid sequence <SEQ ID 4388>. This protein is predicted to be methionine aminopeptidase (map). Analysis of this protein sequence reveals the following:

Possible site: 14

15

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2342(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9761> which encodes amino acid sequence <SEQ ID 9762> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25

>GP:AAC35914 GB:AF071085 methionine aminopeptidase A. [Enterococcus  
 faecalis]  
 Identities = 101/207 (48%), Positives = 128/207 (61%), Gaps = 31/207 (14%)

30

Query: 1 MITLKSAREIEAMDRAQDFLASIHIGLRDIIKPGVDMWEVEEYVRRRCKEENVLPLQIGV 60  
 MITLKS REIE MD +G+ LA +H LR IKPG+ W++E +VR + + QIG  
 Sbjct: 1 MITLKS PREIEMMDES GELLADVHRHLRTFIKPGITSWDIEVFVRDFIESHGGVAAQIGY 60

35

Query: 61 DGAVMDYPYATCCGLNDEVAHAFPRHYTLKQGDLLKVDMLSEPLDKSIVDVSSLNFDNV 120  
 +G Y YATCC +NDE+ H FPR LK GDL+KVDM +  
 Sbjct: 61 EG----YKYATCCSINDEICHGFPRKVKLKDGLIKVDMCVD----- 98

40

Query: 121 AQMKKYTETYSGGLADSCWAYAVGEVSQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180  
 G ++DSCW+Y VGE + E+ LM VT++A+Y+GIE+A +GNRIGDIG A  
 Sbjct: 99 -----LKGALSDSCWSYVVGESTPEIDRLMEVTKKALYLGIEQAQVGNRIGDIGHA 149

Query: 181 IQDYAESRGYGVVRDLVGHGVGPTMHE 207  
 IQ Y E GYGVVRD VGHG+GPT+HE  
 Sbjct: 150 IQTYVEGEGYGVVRDFVGHGIGPTIHE 176

45

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4389> which encodes the amino acid sequence <SEQ ID 4390>. Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

50

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2082(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below.

-1577-

Identities = 256/286 (89%), Positives = 273/286 (94%)

Query: 1 MITLKSAREIEAMDRAAGDFLASIHIGLRDIKPGVDMWEVEEYVRRRCKEENVLPQIGV 60  
 Sbjct: 1 MITLKSAREIEAMDRAAGDFLAGIHIGLRDIKPGVDMWEVEAYVRRRCKEDNVLPQIGV 60

Query: 61 DGAVMDYPYATCCGLNDEVAHAFFRHYTLKQGDLLKVDMLSEPLDKSIVDVSSLNFDNV 120  
 DG +MDYPYATCCGLNDEVAHAFFRHY LK+GDLLKVDMLSEPLDKSIVDV++L+FDNV  
 Sbjct: 61 DGHMMDYPYATCCGLNDEVAHAFFRHYILKEGDLLKVDMLSEPLDKSIVDVAALDFDNV 120

Query: 121 AQMKKYTETYSGGLADSCWAYAVGEVSQEVKDLMSVTREAMYIGIEKAVIGNRIGDIGAA 180  
 +MKK+T +Y+GGLADSCWAYAVG S E+K LM VT+EAMY GIEKAVIGNRIGDIGAA  
 Sbjct: 121 PEMKKWTGSGYTGGLADSCWAYAVGTPSDEIKQLMDVTKEAMYRGIEKAVIGNRIGDIGAA 180

Query: 181 IQDYAESRGYGVVVDLVGHGVGPTMHEEPMVFNYPNYGTAGRGLRLREGMVLTIPEMINTGTW 240  
 +Q+YAES GYGVVVDLVGHGVGPTMHEEPMVFNYPNYGTAGRGLRL+EGMVLTI+PEMINTGTW  
 Sbjct: 181 VQEYAESFGYGVVVDLVGHGVGPTMHEEPMVFNYPNYGTAGRGLRLKEGMVLTIPEMINTGTW 240

Query: 241 EIDTDMKTGWAHKTLDGGLSCQYEHQFVITKDGPIVILTSQGEERTY 286  
 EIDTD+KTGWAHKTLDGGLSCQYEHQFVITKDGPIVILTSQGEERTY  
 Sbjct: 241 EIDTDIKTGAHKTLDGGLSCQYEHQFVITKDGPIVILTSQGEERTY 286

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1430

A DNA sequence (GBSx1516) was identified in *S.agalactiae* <SEQ ID 4391> which encodes the amino acid sequence <SEQ ID 4392>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3473 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9759> which encodes amino acid sequence <SEQ ID 9760> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06894 GB:AP001518 unknown conserved protein [Bacillus halodurans]  
 Identities = 158/431 (36%), Positives = 270/431 (61%), Gaps = 6/431 (1%)

Query: 6 SKHQEILEYLENLAVGKRVSVRSISNHLKVS DGTAYRAIKEAENRGIVETRPRSGTVRVA 65  
 +KH++IL+Y+ NL VG+++SVR I+ L+VS+GTAYRAIKEAEN+G+V T R GT+R+  
 Sbjct: 3 TKHEQILQYITNLEVGEKISVRRIAKDLQVSEGTAYRAIKEAENQGLVSTIERVGTIRIE 62

Query: 66 QKAKVNIKLTLYAEIARISDSQVVAGIEGLSKEFSKFSIGAMTHRNIEKYLVOGGLLIVG 125  
 +K K NIEKLTLYAE+ I D QV+ G +GL K ++F IGAM + +Y+ G LLIVG  
 Sbjct: 63 KKQKENIEKLTLYAEVNVNIVDQVGLGGRDGLHKTILNRFVIGAMKLDAMMRYVEPGNLLIVG 122

Query: 126 DRDEIQHLALQHQNAILVTGGFNVSPVCRLADKLQIPVMVTHYDTFTVSTMINHTLSNA 185  
 +R ++ +AL+ A+L+TGGF+ S +LAD+L +PV+ T YDTFTV+TMIN + +  
 Sbjct: 123 NRYQVHQIALEAGAAVLITGGFDTSDAIAKLADELDPVISTSYDTFTVATMINRAIYDQ 182

Query: 186 KIRTDLKTVEQVYQSQMDYGFQAQDDTVKEFNLLVKQTKNVRFPVFNQANVVGVVSVQD 245  
 I+ ++ V+ + D ++ ++ V +++ L ++T + R+P++++ + G+V+ +D  
 Sbjct: 183 LIKKEITLVDDILIPLDQTYMTTENNVVGKWHELNEKTGHSRYPVIDENMKIQGMVAARD 242

Query: 246 ILGKDKEVKLATVMSKNIIVAKPRMSLANISQKMFEDLNMPVVSDDFELLGVITRRQA 305  
 +L + + VM+KN I R S+A ++ M++E + ++PV+ +L+GV++R+  
 Sbjct: 243 VLNASRHTPIEKVMTKNPITVSERTSVAVAHVMMWEGIELLPVIDSHRKLIGVVSQRDV 302

-1578-

Query: 306 VENLSMSQ-----GTDLYTYSQDILSNLQIEDG-HFSFLVEPAMIDHTGSLTQGVLTFL 359  
 ++ L M Q G + L+ + G + + P M + G+++ GV+T +  
 Sbjet: 303 LKALQMIQRQPHVGETIEDLMTNGLNESSDQGDSYEVEITPQMTNQLGTISHGVMTSLV 362

Query: 360 KEICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVPTTIISEKRREATLDLELKLENK 419  
 E RVL + + +VV+ +TLYFL+PVQID + + P ++ R+ +D+E+ E +  
 Sbjet: 363 IESGSRVLRKYKKGDLVVENITLYFLKPVQIDSRLTIRPRVLEIGRKHGKIDVEMYHEGE 422

Query: 420 IIAKAMIAVKI 430  
 I+ KA+ +I  
 Sbjet: 423 IVGKALFMAQI 433

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4393> which encodes the amino acid  
 sequence <SEQ ID 4394>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3011(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 267/431 (61%), Positives = 351/431 (80%)

Query: 1 MIIVMSKHQEIILEYLENLAVGKRVSVRSISNHLKVSDGTAYRAIKEAENRGIVETRPRSG 60  
 +II+MSKHQ+IL+YLE LA+GK+VSVRSISNHLKVSDGTAYRAIKEAENRGIVET+PRSG  
 Sbjet: 1 VIIIMSKHQDILDYLEKLAIGKKVSVRSISNHLKVSDGTAYRAIKEAENRGIVETKPRSG 60

Query: 61 TVRVAQKAKVNIEKLTYAEIARISDSQVVAGIEGLSKEFSKFSIGAMTHRNIEKYLQVGG 120  
 TVR+ +K +V I++LTY+EIARISDS+V+AG GL EFS+FSIGAMT +NI +YLV+GG  
 Sbjet: 61 TVRIEKKGRVRIDRLTYSEIARISDSSEVLGHAGLGHEFSRFSIGAMTQQNIRRYLVKGG 120

Query: 121 LLIVGDRDEIQHLALQHONAILVTGGFNVSPVCRLADKLQIPVMVTHYDTFTVSTMINH 180  
 LLIVGDR+ IQ LAL++ NAILVTGGF VS V +A+ +IPVMVTHYDTFTV+TMINH  
 Sbjet: 121 LLIVGDRETIQLLALEHNHAILVTGGFPVSKRVIEMANNQRIPVMVTHYDTFTVATMINH 180

Query: 181 TLSNAKIRTDLKTVEQVYQSQMDYGFQAQDDTVKEFNLLVKQTKNVRFPVQNQANVVVGV 240  
 LSN +I+TDLKTVEQV DYGL +D +V+EFN L+K+T+ VRFP+++ V+GV  
 Sbjet: 181 ALSNIRIKTDLKTVEQVMIPITDYGLCEDSSVEEFNTLIKKTRQVRFPVLDYKRKVIGV 240

Query: 241 VSVQDILGKDKEVKLATVMSKNIIIVAKPRMSLANISQKMIFEDLNMPVVSDDFELLGVI 300  
 VS++D++ + KL VMSKN I A+P SLANISQKMIFEDLNMPV++ LLG+I  
 Sbjet: 241 VSMRDVVDQLPTTKLTKVMSKNPITARPNTSLANISQKMIFEDLNMLPVTDEENNLLGMI 300

Query: 301 TRRQAVENLSMSQGTDLTYTSDQILSNLQIEDGHFSFLVEPAMIDHTGSLTQGVLTFLK 360  
 TRRQA+ENL Q + YTYS+QILSNL+ ++ +VEP MID G+++ GV++EFLK  
 Sbjet: 301 TRRQAMENLPNHQPNNPYTYSEQILSNLEETVDYQVVEPTMIDSAGNMSGVISEFLK 360

Query: 361 EICIRVLTRKHQRSIVVKQMTLYFLQPVQIDEIIMVPTTIISEKRREATLDLELKLENKI 420  
 EI IR LT+KHQ++I+++QM +YFL +QI++ + + P II+E RR +T+D+E+ +++++  
 Sbjet: 361 EISIRALTKKHQKNIIIEQMMVYFLHAIQIEDELKIYPKIITENRRSSTIDIEIFVDDQV 420

Query: 421 IAKAMIAVKIN 431  
 IAKA+I KIN  
 Sbjet: 421 IAKAIITTKIN 431

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

-1579-

**Example 1431**

A DNA sequence (GBSx1517) was identified in *S.agalactiae* <SEQ ID 4395> which encodes the amino acid sequence <SEQ ID 4396>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2837(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]

Identities = 56/185 (30%), Positives = 86/185 (46%), Gaps = 4/185 (2%)

Query: 7 MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFI-FPVQTSKKESDFLLV 65

M+I G +ETE + LR D A + AS +++ + S K+S+ L

Sbjct: 1 MEIEDIYGDLPTELETERLRKFKYKDDAAAIYDYASNEQVTKYVLWETHQSIKDSEAFLLA 60

Query: 66 HSPMK---EPLGVWAIEDKVSHKMGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECL 122

+ K + + WAIE K + +M G + F KTAE+GY L E WQGIMTE +

Sbjct: 61 FALNKYDEKDVSPWAIELKRNERMIGTVDFVWWKPKDKTAEGLGYVLSEPYWGQGIMTEAV 120

Query: 123 KTLFFFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDDRYTRRIRDYIEFQL 182

L F F +++++ ENI+S +V KA + + + RD+ + +

Sbjct: 121 NALVEFGFNNMELERIQAKCFAENISSARVMEKAGLIYEGTHRRAIYVKGHRDFKVYAI 180

Query: 183 TRGDY 187

R DY

Sbjct: 181 IREDY 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 667> which encodes the amino acid sequence <SEQ ID 668>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1096(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 94/177 (53%), Positives = 117/177 (65%)

Query: 7 MDIWTNLGRFAFIETEHVNLRPVAYTDREAFWRIASKRTNLQFIFPVQTSKKESDFLLVH 66

MDIWT L FAF ET V LRP Y D F+ + + NL ++FP Q +K SD+LLVH

Sbjct: 1 MDIWTKLAVFAFFETPKVILRPFRYEDHWDYFYSMVNDTKNLYYVFPEQKTKAASDYLLVH 60

Query: 67 SFMKEPLGVWAIEDKVSHKMGVIRFENIDLSKKTAEIGYFLKESSWGQGIMTECLKTL 126

SF+K PLG WAIEDK +H++ G IR E+ D + A+IGYFL + WQGIMTE + L

Sbjct: 61 SFIKFPFGQWAIEDKATHQVIGSIRIEHYDAKTRCADIGYFLNYAFWQGIMTEVVIKLV 120

Query: 127 FFAFREFGMDKLIIVTHKENIASQKVALKAHFKQSRSFKGSDDRYTRRIRDYIEFQLT 183

+ +F EFG+ L I+TH EN ASQKVA KA F+ FKGSDR T +I Y +QLT

Sbjct: 121 YLSFHEFGLKTLRIITHLENKASQKVAKAGFQLKTCFKGSDRNTHKICIKYMYQLT 177

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1580-

**Example 1432**

A DNA sequence (GBSx1518) was identified in *S.agalactiae* <SEQ ID 4397> which encodes the amino acid sequence <SEQ ID 4398>. This protein is predicted to be UDP-N-acetylglucosamine-1-carboxyvinyl transferase (murA). Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -5.63    Transmembrane    25 - 41 ( 24 - 42)

    ----- Final Results -----
10      bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 240/412 (58%), Positives = 303/412 (73%), Gaps = 2/412 (0%)

20  Query: 3   KIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVLDGVPALSDVDSLVDIMETMGA 62
      K+II GGK+L G + V GAKNS VALIPA ILa+ VVL+G+P ISDV +L +I+E +G
      Sbjct: 20 KLIIRGGKKLAGTLQVDGAKNSAVALIPAAILAESEVVLEGLPDISDVHTLYNILEELGG 79

      Query: 63 KIKRYGETLEIDPCGVKDIPMPYKINSLRASYYFYGSLLGRYQATLGLPGGCDLGPRP 122
      ++ +T IDP + +P+P G + LRASY G++LGR+ +A +GLPGGC LGPRP
25  Sbjct: 80 TVRYDNKTAVIDPTDMISMPLESGNVKKLRASYLGMAGLGRFKAIVIGLPGGCYLGPRP 139

      Query: 123 IDLHLKAFEAMGASVSIEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAKANGRT 182
      ID H+K FEA+GA V+ E ++ L + L+GA IY+D VSVGATIN ++AA +A G+T
30  Sbjct: 140 IDQHIKGFEALGAKVTNEQGAIYLRAD--ELKGARIYLDVSVGATINIMLA AVRAGKT 197

      Query: 183 VIENAAEPEIIDVATLLNNMGAIHAGTDTVITIEGVKSLHGTRHQVIPDRIEAGTYIA 242
      VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG H +IPDRIEAGT++
      Sbjct: 198 VIENAAEPEIIDVATLLNMGAIIKGAGTDTIRITGVEHLHGCHHTIIPDRIEAGTFMV 257

35  Query: 243 MAAAIGRGIKVTNVLYEHLESFIAKLDEMGVVMTVEEDSIFVEEQERLKAVSIKTSPPYG 302
      +AAA G+G+++ NV+ HLE IAKL EMGV M +EED+IFV E E++K V IKT YPG
      Sbjct: 258 LAAASGKGVRIENVIPHLEGIIAKLTEMGVVPMDEEDAI FVGEVEKIKKVDIKTYAYPG 317

      Query: 303 FATDLQQPLTPLLTLTAEGNGSLDDTIYEKRVNHVPELARMGANISTLGGKIVYSGPNQLS 362
      F TDLQQPLT LL AEG+ + DTIY R H+ E+ RMG G V +GP QL
40  Sbjct: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAEIERMGGKFKLEGRSAVINGPVLQ 377

      Query: 363 GAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIEKLTSLGADI 414
      G+ V ATDLRAGAALVIA L+A+G TEI +E I RGYS IIEKL+++GA+I
45  Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIIEKLSAIGANI 429

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4399> which encodes the amino acid sequence <SEQ ID 4400>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 21
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -8.70    Transmembrane    25 - 41 ( 23 - 45)

    ----- Final Results -----
55      bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

60  >GP:AAF86297 GB:AF072894 UDP-N-acetylglucosamine-1-carboxyvinyl
      transferase [Listeria monocytogenes]
      Identities = 244/412 (59%), Positives = 302/412 (73%), Gaps = 2/412 (0%)

```

-1581-

Query: 3 KIIINGGKALSGEVAVSGAKNSVVALIPAILADDIVILDGVPASDVDSLIEIMELMGA 62

5 Sbjct: 20 KLIIRGGKKLAGTLOVDGAKNSAVALIPAAILAESEVVLEGLPDISDVHTLYNILEELGG 79

Query: 63 TVNYHGDITLEIDPRGVQDIPMPYKINSLRASYYFYGSLLGRFGQAVVGLPGGCDLGRP 122  
TV Y T IDP + +P+P G + LRASY G++LGRF +AV+GLPGGC LGPRP

10 Sbjct: 80 TVRYDNKTAVIDPTDMISMPPLSGNVKKLRASYILMGAMLGRFCKAVIGLPGGCYLGRP 139

Query: 123 IDLHLKAFEAMGVEVS YEGENMNLSTNGQKIHGAHIYMDTVSVGATINTMVAATKAQGKT 182  
ID H+K FEA+G +V+ E + L + ++ GA IY+D VSVGATIN M+AA +A+GKT

15 Sbjct: 140 IDQHIKGFALGAKVTNEQGAITYLRAD--ELKGARIYLDVVSVGATINIMLA AAVRAKGKT 197

Query: 183 VIENAAAREPEIIDVATLLNNMGAHIRGAGTDIITIQQVQKLHGTRHQVIPDRIEAGTYIA 242  
VIENAA+EPEIIDVATLL NMGA I+GAGTD I I GV+ LHG H +IPDRIEAGT++

20 Sbjct: 198 VIENAAAREPEIIDVATLLNNMGAIKAGAGTDITIRITGVEHLHGCHHTIIPDRIEAGTFMV 257

Query: 243 LAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAI FVEKQESLKAITIKTSPYPG 302  
LAAA GKGV+I NV+ HLE IAKL EMGV M +EEDAI FV + E +K + IKT YPG

25 Sbjct: 258 LAAASGKGVRIENVIPHTLEGIIAKLTEMGVPMDEEDAI FVGEVEKIKKVDIKTYAYPG 317

Query: 303 FATDLQQPLTPLLKADGRGTIIDTIYEKRINHVPPELMRMGADISVIGGQIVYQGPPSRLT 362  
F TDLQQPLT LL +A+G I DTIY R H+ E+ RMG + G V GP +L

30 Sbjct: 318 FPTDLQQPLTALLTRAEGSSVITDTIYPSRFKHIAEIERMGGKFKLEGRSAVINGPVQLQ 377

Query: 363 GAQVKATDLRAGAALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADI 414  
G++V ATDLRAGAALV A L+A+G+TEI +E I RGY+ II KL+A+GA+I

30 Sbjct: 378 GSKVTATDLRAGAALVIAALLADGETEIHGVEHIERGYSKIIEKLSAIGANI 429

An alignment of the GAS and GBS proteins is shown below.

Identities = 344/419 (82%), Positives = 394/419 (93%)

35 Query: 1 MRKIIINGGKQLTGEVAVSGAKNSVVALIPATILADDVVVLDGVPASDVDSLVDIMETM 60  
MRKIIINGGK L+GEVAVSGAKNSVVALIPA ILADD+V+LDGVPASDVDSL++IME M

Sbjct: 1 MRKIIINGGKALSGEVAVSGAKNSVVALIPAILADDIVILDGVPASDVDSLIEIMELM 60

40 Query: 61 GAKIKRYGETLEIDPCGVKDIPMPYKINSLRASYYFYGSLLGRYQATLGLPGGCDLGP 120  
GA + +G+TLEIDP GV+DIPMPYKINSLRASYYFYGSLLGR+GQA +GLPGGCDLGP

Sbjct: 61 GATVNYHGDITLEIDPRGVQDIPMPYKINSLRASYYFYGSLLGRFGQAVVGLPGGCDLGP 120

45 Query: 121 RPIDLHLKAFEAMGASVSYEGDSMRLATNGKPLQGANIYMDTVSVGATINTIIAAKANG 180  
RPIDLHLKAFEAMG VSYEG++M L+TNG+ + GA+IYMDTVSVGATINT++AA KA G

Sbjct: 121 RPIDLHLKAFEAMGVEVS YEGENMNLSTNGQKIHGAHIYMDTVSVGATINTMVAATKAQG 180

50 Query: 181 RTVIENAAAREPEIIDVATLLNNMGAHIRGAGTDVITIEGVKSLHGTRHQVIPDRIEAGTY 240  
+TVIENAAAREPEIIDVATLLNNMGAHIRGAGTD+ITI+GV+ LHGTRHQVIPDRIEAGTY

Sbjct: 181 KTVIENAAAREPEIIDVATLLNNMGAHIRGAGTDIITIQQVQKLHGTRHQVIPDRIEAGTY 240

55 Query: 241 IAAAAIGRGKIKVTNVLYEHLESFIAKLDEMVRMTVEEDSIFVEEQERLKAVSIKTSPI 300  
IA+AAAIG+G+K+TNVLYEHLESFIAKL+EMGVRMTVEED+IFVE+QE LKA++IKTSPI

Sbjct: 241 IALAAAIGKGVKITNVLYEHLESFIAKLEEMGVRMTVEEDAI FVEKQESLKAITIKTSPI 300

60 Query: 301 PGFATDLQQPLTPLLTAEGNGSLDITIYEKRVNHVPELARMGANISTLGGKIVYSGPNQ 360  
PGFATDLQQPLTPLL A+G G++DTIYEKRVNHVPEL RMGA+IS +GG+IVY GP++

Sbjct: 301 PGFATDLQQPLTPLLKADGRGTIIDTIYEKRINHVPPELMRMGADISVIGGQIVYQGPPSR 360

Query: 361 LSGAPVKATDLRAGAALVIAGLMAEGRTEITNIEFILRGYSNIEKLTSLGADIQLVEE 419  
L+GA VKATDLRAGAALV AGL+AEG+TEITNIEFILRGY++II KLT+LGADIQL+E+

60 Sbjct: 361 LTGAQVKATDLRAGAALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADIQLIED 419

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1582-

**Example 1433**

A DNA sequence (GBSx1519) was identified in *S.agalactiae* <SEQ ID 4401> which encodes the amino acid sequence <SEQ ID 4402>. This protein is predicted to be thiamine phosphate pyrophosphorylase (thiE). Analysis of this protein sequence reveals the following:

```

5   Possible site: 55
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.0422(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:AAF25544 GB:AF109218 ThiE [Staphylococcus carnosus]
      Identities = 98/200 (49%), Positives = 140/200 (70%), Gaps = 1/200 (0%)

      Query: 5   LKLYFVCGTVDCSR-KNILTVEEALQAGITLQFREKGFTALQGKEKIAMAKQLQILCK 63
                L +YF+CGT D   + I  V++EAL+ GITL+QFREKG A  G++K+A+AK+LQ LCK
20   Sbjct: 7   LNVYFICGTQDIPEGRTIQEVLKEALEGGITLYQFREKNGAKTGQDKVALAKELQALCK 66

      Query: 64  QYQVPFIIDDDIDLVELIDADGLHIGQNDLPVDEARRRLPDKIIGLSVSTMDEYQKSQLS 123
                Y VPFI++DD+ L E IDADG+H+GQ+D VD+   R  KIIGLS+ ++E  S L+
      Sbjct: 67  SYNVPFIVNDDVALAEEDADGIHVGQDDEAVDDFNNRFEGKIIGLSIGNLEELNASDLT 126

25   Query: 124  VVDYIGIGPFNPQTQSKADAKPAVGNRITTKAVREINQDIPIVAIGGITSDFVHDIIESGAD 183
                VDYIG+GP  T SK DA   VG +  + +R+  D+PIVAIGGI+ D V ++ ++ AD
      Sbjct: 127  YVDYIGVGPIFATPSKDDASEPVGPKMIETLRKEVGDLPIVAIGGISLDNVQEVAKTSAD 186

30   Query: 184  GIAVISAIKANHIVDATRQ 203
                G++VISAI+++ H+ +  +
      Sbjct: 187  GVSVISAIARSPHVTETVHK 206

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1434**

A DNA sequence (GBSx1520) was identified in *S.agalactiae* <SEQ ID 4403> which encodes the amino acid sequence <SEQ ID 4404>. This protein is predicted to be hydroxyethylthiazole kinase (b2104). Analysis of this protein sequence reveals the following:

```

40   Possible site: 54
   >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -4.94    Transmembrane 198 - 214 ( 194 - 217)

   ----- Final Results -----
45          bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8805> which encodes amino acid sequence <SEQ ID 8806> was also identified. Analysis of this protein sequence reveals the following:

```

55   Lipop: Possible site: -1    Crend: 7
      McG: Discrim Score:      -2.93
      GvH: Signal Score (-7.5): 1.61
      Possible site: 39
   >>> Seems to have no N-terminal signal sequence

```

-1583-

ALOM program count: 1 value: -4.94 threshold: 0.0  
 INTEGRAL Likelihood = -4.94 Transmembrane 183 - 199 ( 179 - 202)  
 PERIPHERAL Likelihood = 2.49 151  
 modified ALOM score: 1.49

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.2975 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF25543 GB:AF109218 ThiM [Staphylococcus carnosus]  
 Identities = 114/253 (45%), Positives = 160/253 (63%), Gaps = 1/253 (0%)

Query: 18 LEQLKEVNPLTICITNNVKNFTANGLLALGASPAMSECIEDLEDLLKVADALLINIGTL 77  
 L+Q++ +PL IC TN+VVKNF+TANGLL+LGASP MSE ++ ED VA ++LINIGTL  
 Sbjct: 5 LDQIRTEHPLVICYTNDVVKNF+TANGLLSLGASPTMSEAPQEAEDFYFVAGSVLINIGTL 64

Query: 78 TKESWQLYQEAIKIANKNQVPVVLDPVAAGASRFRLEVSLDLLKNYSISLLTNGNSEIAA 137  
 TK E KIAN+ + P+V DPVA GAS++R + LK +++ GN SEI A  
 Sbjct: 65 TKHHEHAMLENAKIANETETPLVFDPAVVGASKYRKDFCKYFLKKIKPTVIKGNASEILA 124

Query: 138 LIGKQASKGADGGKVADLESIAVKANQVDFVPPVVTGETDAIAVRGEVRLQNGSPLMP 197  
 LI + KG D D+ IA KA ++ ++TGETD I +V L NGS +  
 Sbjct: 125 LIDDTATMKGTDSDNLDVVDIAEKAYKEYQTAILTGETDVIVQDNKVVKLSNGSHFLA 184

Query: 198 LVTGTGCLLGAVLAFIGSSDRSDDLACLTEAMTVYNVAGEIAEKVAKGKGVGSFQVAF 257  
 +TG GCLLGAV+ AF+ + + L EA++VYN+A E AE+++ KG G+F F+  
 Sbjct: 185 KITGAGCLLGAVVGAFI-FRNTHPSIETLIEAVSVYNIAERAELSDSKGPGTFLTQFI 243

Query: 258 DALSQMKSEMIMD 270  
 DAL ++ S+ + +  
 Sbjct: 244 DALYRIDSDAVAE 256

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8806 (GBS398) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 6; MW 31.8kDa).

The GBS398-His fusion product was purified (Figure 214, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 314), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1435

A DNA sequence (GBSx1521) was identified in *S.agalactiae* <SEQ ID 4405> which encodes the amino acid sequence <SEQ ID 4406>. This protein is predicted to be ThiD (thiD). Analysis of this protein sequence reveals the following:

Possible site: 44  
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1584-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25542 GB:AF109218 ThiD [Staphylococcus carnosus]
Identities = 139/258 (53%), Positives = 186/258 (71%), Gaps = 4/258 (1%)

5   Query: 8   LTIAGTDPGGAGIMADLKTFFQARRTYGMAVVTSVVAQNTCGVRGVQHIETAIIDQQLAC 67
      LTIAGTDP+GGAG+MADLK+F A   YGMA +TS+VAQNT GV+ + +++ + +QL
      Sbjct: 8   LTIAGTDPTGGAGVMADLKSFHACGVYGMAAITSIVAQNTKGVQHIHNLDTWLKEQLDS 67

10  Query: 68  VYDDIKPKAVKTGMLAERETISLVASYLKKYPQ-PYVLDPVMVATSGHRLIDSDAVEALK 126
      ++DD P+A+KTGM+A +E + L+ SYL+KYP PYV+DPVM+A SG L+D AL+
      Sbjct: 68  IFDDELFPQAIKTGMIATKEMMELIRSYLEKYPDIPYVIDPVMLAKSGDSLMDAGKHALQ 127

      Query: 127 EDLLPLATIITPNLPEAEVLVG YDLSDEVSIKAGYDIQKQYSVRNVLIKGGHLD--GLA 184
      E LLPLA + TPNLPEAE +VG+ L E +I KAG + + V+IKGGH++ +A
15  Sbjct: 128 EILLPLADVATPNLPEABEIVGFKLDTEEAIKKAGDIFINEIGSKGVVVIKGGHIEDKNIA 187

      Query: 185 KDYLFLEKEGLITLSNQRINTIHTHTGCTFAAVVAAELAKGQSILNAVSTAKSFITS AI 244
      KDYLF K+GL ++R +T HTHGTGCTF+AV+ AELAKG++I AV AK FI +I
20  Sbjct: 188 KDYLF-TKDGLEVFESERYDTKHTHTGCTFSAVITAEALAKGKTIYEAVKKAKDFIALSI 246

      Query: 245 ETAPELGLGNGP VNH TS Y 262
      + PE+G G G P VNH +Y
      Sbjct: 247 KYTPEIGQGRGP VNH FAY 264
```

25 There is also homology to SEQ ID 4408.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1436

A DNA sequence (GBSx1522) was identified in *S.agalactiae* <SEQ ID 4409> which encodes the amino acid sequence <SEQ ID 4410>. This protein is predicted to be TenA (tenA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
```

35 ----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2242(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

40 The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAF25541 GB:AF109218 TenA [Staphylococcus carnosus]
Identities = 78/213 (36%), Positives = 127/213 (59%), Gaps = 6/213 (2%)

45  Query: 14  IQSIYQDPFIQGIKGRILDHDVICHYLOADNIYLGKFADIYALCLAKSDNLRDKQFFLEQ 73
      I IYQD FIQ ++KG + + + YL+AD YL +FA+IYAL + +L +F ++Q
      Sbjct: 15  IDEIYQDHFIEQLLKGDIKKEALRQYL RADASYLREFAN IYALLIPIMPDLESVRFLVDQ 74

      Query: 74  IDFTLNRELADGEGPHQALAA YTNRSYQDIIEKGVWYPSADHYIKHMYFHFY-ENGIAGA 132
      I F +N E+ H+ +A Y +Y +I++K VW PS DHYIKHMY++ Y A A
50  Sbjct: 75  IQFIVNGEVE----AHEYMADYIGENYNEIVQKKVWPPSGDHYIKHMYYNVYAHENAAYA 130

      Query: 133 LAAMSPCPWIYHQLAKKII EENQFLNGNPFNNWITFYANDTVEELMENYFRMMDYYAQN L 192
      +AAM+PCP++Y +AK+ +++ + W FY N ++ L+E +M+ N+
55  Sbjct: 131 IAAMAPCPVYVYAMIAKRAMKDPNLNKSSILAKWFEFY-NTEMDPLIEVLDDLMNQLTANM 189

      Query: 193 SKEKQADLVDAFVKSCQHERRFFQMAINQEKWE 225
      S+ ++ ++ + +++S HE FF MA EKW+
      Sbjct: 190 SETEKNEVRENYLQSTVHELNFNNMAYTSEKWQ 222
```

-1585-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1437

- 5 A DNA sequence (GBSx1523) was identified in *S.agalactiae* <SEQ ID 4411> which encodes the amino acid sequence <SEQ ID 4412>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have a cleavable N-term signal seq.

10 INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 ( 36 - 63)  
 INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 ( 92 - 112)  
 INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 ( 135 - 151)  
 INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 ( 69 - 85)  
 INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 ( 216 - 232)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91230 GB:Z56283 orf2 [Lactobacillus helveticus]

Identities = 46/215 (21%), Positives = 96/215 (44%), Gaps = 3/215 (1%)

25 Query: 21 AITFLCLLIPTFSFSFTLRRLTSLFLIIVVTLCQFVKVSLKTWAKVNLISFVMGLSLFL 80  
 ++ F+ I + S L T+L+ + + ++ +K + + F+ ++F  
 Sbjct: 4 SLKFILAFIISLEISLKASLTNLIIVIAFALIYLLVTRIKIKELILLIIVPFIASFTIFA 63

30 Query: 81 GTYFWGKLPHQFVLASLVACRPLIFMNVGLLFHASHSNYDFIESLYQTFKVP SHFAYGIF 140  
 +++ P + +L + R ++ + + DF SL Q +PS FAYG+  
 Sbjct: 64 TLFWFSPTPDAYYAWN L-STRVYVYTLTIACVTRNTTATDFARSLEQNLHLPSKFAYGV L 122

35 Query: 141 AVFNLLPLIKLQYQRNRLAFLRKNQVTWALSPRLILSVLLKTIYWVEQLELAML SKGFEV 200  
 A N++P +K ++ R + ++ SP L +L + + L M S G+  
 Sbjct: 123 AAINIIPRMKTAVKQIRTSAMMRGMYLSFWSPVLYFKAILVALNSADNLAQGMESHGYVE 182

Query: 201 GKERTHASTYPVRFRDYS L-LGMSILLSIGM-IFK 233  
 G++R P+ +D+ + + IL++I + IFK  
 Sbjct: 183 GQKRATIVAIP LTKDWLIFFTLLILVNISLFIK 217

- 40 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8807> and protein <SEQ ID 8808> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

McG: Discrim Score: 4.50

45 GvH: Signal Score (-7.5): -0.2

Possible site: 35

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 5 value: -7.06 threshold: 0.0

50 INTEGRAL Likelihood = -7.06 Transmembrane 43 - 59 ( 36 - 63)  
 INTEGRAL Likelihood = -2.55 Transmembrane 92 - 108 ( 92 - 112)  
 INTEGRAL Likelihood = -1.49 Transmembrane 135 - 151 ( 135 - 151)  
 INTEGRAL Likelihood = -1.06 Transmembrane 69 - 85 ( 69 - 85)  
 INTEGRAL Likelihood = -0.22 Transmembrane 216 - 232 ( 216 - 232)  
 PERIPHERAL Likelihood = 2.65 170  
 55 modified ALOM score: 1.91

\*\*\* Reasoning Step: 3

-1586-

----- Final Results -----

bacterial membrane --- Certainty=0.3824(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1438**

A DNA sequence (GBSx1524) was identified in *S.agalactiae* <SEQ ID 4413> which encodes the amino acid sequence <SEQ ID 4414>. Analysis of this protein sequence reveals the following:

10

Possible site: 24

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

15

bacterial cytoplasm --- Certainty=0.3007(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20

&gt;GP:CAA91229 GB:Z56283 orf1 [Lactobacillus helveticus]

Identities = 123/424 (29%), Positives = 200/424 (47%), Gaps = 48/424 (11%)

25

Query: 17 LFDEVTFSLNPGERILISGYSGCGKSTLALLLSGL--KESGK--GQVLLNGSLIEPSDVG 72  
 L +++ ++ PG +LI G +GCGKSTL +++GL K +GK G++ L+G  
 Sbjct: 12 LINQLNMNIAPGFNLLI-GPTGCGKSTLLKIAGLYPKYAGKLTGKIDLHGQ-----KAA 65

30

Query: 73 FLFQNPDLQFCMDTVAHELYFILENLQIEPEQMQRSEFVLAQVGLKGFQNRILIYTLSSQG 132  
 +FQN QF M T E+ F LENLQI+ + + + + ++ I TLS G  
 Sbjct: 66 MMFQNAAEQFTMTTPREIIFALENLQIKAKDYDLHIKKAFFTKIADLLDQKINTLSGG 125

35

Query: 133 EKQRLALATIFLKSPKLIILDEAFANLDQESASQLLQVLNYQANNQSMILVIDHLITYY 192  
 ++Q +ALA + + +LDE FA+ D + L++ + + ++ +I+ DH++ Y  
 Sbjct: 126 QQQHVALAVLIAMDVDVFLDDEFFASCDPNTRHFLIEKLASLAETGRT-IILSDHVLDDY 184

40

Query: 193 QDIMDHYFWLEKRLTRVNFDMNRLNVFELEKKSHN-----TGDKLLSIKDFQVK- 243  
 + I DH + E + + N+L F+ K+ H TG + + Q+K  
 Sbjct: 185 EKICDHLQYFEGKTVKELSANENKKL--FKQNKQFHEQSYFALPTGTPVFELNKTQIKQ 242

45

Query: 244 ----LSKNKFISYLDLDFDLASGERLCLDGPSSGVGKSSLFMGLLGLYRTKGK-----KQ 291  
 L +NK Y G+ + G +GVGK+SLF + + KG +  
 Sbjct: 243 NRLLLKQNKLLKIY-----GKTTLITGSGNGVGTSLFKAMTKMIPYKGNFTYLDNEISK 295

50

Query: 292 FTHRKQIP-ISFLFQNPDLQFIFSTVYDEIFQVCKDSN-----KARDILETINLWDKKQ 344  
 +RK + I+ FQ DQF+ TV DEI KD N K + LE + L  
 Sbjct: 296 IKYRKYLSQIAQFFQKASDQFLTFTVKDEIELSKDRNNFFTDKIDEWLEKLQKQHL 355

Query: 345 FSPFQLSQCGQRRILAIGSILASDSKLLLLDEPTYGQDAYHANMITTLSSYCHKNHCGVI 404  
 + LS GQQ++L I +L + +LL+DEP G D +++ L+ K +  
 Sbjct: 356 QVVYSLSGGQKKLQILMLMTKHNVLIDEPLSGLDHESVDLVLQLMQECQEKLQQTFL 415

Query: 405 FTSH 408  
 SH

Sbjct: 416 IISH 419

55

Identities = 44/185 (23%), Positives = 83/185 (44%), Gaps = 24/185 (12%)

Query: 28 GERILISGYSGCGKSTLALLLSGLKESGKGQVLLNGSLIEP-----SDVGFLFQNPDLQ 81  
 G+ LI+G +G GK++L ++ + L+ + + S + FQ Q  
 Sbjct: 256 GKTTLITGSGNGVGTSLFKAMTKMIPYKGNFTYLDNEISKIKYRKYLSQIAQFFQKASDQ 315

60

Query: 82 FCMDTVAHELYFILENLQIEPEQMQRSEFV-----LAQVGLKGFQNRILIYTLSSQGE 133  
 F TV E+ +DR+ F L ++ LK ++++Y+LS G+

-1587-

Sbjct: 316 FLTVTVKDEIEL-----SKKDRNNFFTDKIDEWLEKIQLKQHLDQVVYSLSGGQ 365

Query: 134 KQRLALATIFLKSPKLIILDEAFANLDQESASQLLQVLVNYQANNQSMILVIDHLITYYQ 193  
 +++L + + + ++++DE + LD ES +LQL+ Q Q ++I H I

5 Sbjct: 366 QKKIQILLMLMTKHNVLIDEPLSGLDHESVDLVLQLMQECQEKLOQTFLIISHQIDALA 425

Query: 194 DIMDH 198  
 D D+

10 Sbjct: 426 DFCDY 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4415> which encodes the amino acid sequence <SEQ ID 4416>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3093(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

An alignment of the GAS and GBS proteins is shown below.

Identities = 120/455 (26%), Positives = 203/455 (44%), Gaps = 47/455 (10%)

Query: 1 MLSVEKLACTHGD SHYLFDEV-TFSLNPGERILISGYSGCGKSTLALLLSGLKE---SGK 56  
 M+S E+L T+ D ++ T + G+ I++ G SG GKST LL+G+ +GK

25 Sbjct: 21 MISAEQLVFTYHDQKNPACQISTCQIASGQFIVLCGPGSGKSTFLKLLNGIIPDYAGK 80

Query: 57 GQVLNGLSLIEPS-----DVGFLFQNPDLQFCMDTVAHELYFILENLQIEPEQMMD 107  
 + L+ + + V +FQNP QF V HEL F EN + + + +

30 Sbjct: 81 YEGRLDVADCQAGRDSVETFSRSVASVFQNPASQFFYREVQHELVPFCENQGLDAKVIK 140

Query: 108 RSEFVLAQVGLKGFQNRLIYTLQSGEKQRLALATIFLKSPKLIILDEAFANLDQESASQL 167  
 R + N+ ++ LS G+KQR+A+AT ++ +++ DE ANLD + +

35 Sbjct: 141 RLWTLAEDFAFAELLNKMDFGLSGGQKQKRVATATIMQGTNIMLFDEPTANLDSAGIAAV 200

Query: 168 LQVLVNYQANNQSMILVIDHLITYYQDIMDHYFW-----LEKRLTRVNF-----DY 213  
 + +A ++ +IV +H + Y D+ D++F+ L +LT N D

40 Sbjct: 201 KAYLTQLKAAGKT-IIVAEHRHLHYLMDLADNFFYFKNGRLTDKLTQNLALTDEQRQDM 259

Query: 214 MLNRLNVFELE-----KKSHNTGDKLLSIKDFQVKLSKNKFISYLDLDFLASGERLCLD 266  
 L'RL++ +L+ .. + H D L I+ V+ A G +

45 Sbjct: 260 GLRRLDLSDLKPVLAGKIESQHVRPDDSLCIEHLTVRAGSKILRCIEQLSFVAGSISGIT 319

Query: 267 GPSGVGKSSLFMGLLGLYRTKGKKQFTHRKQIPISFLFQNPDLQFIFSTVYDEIF--QVC 324  
 G +G+GKS L + G+ KK + IP+S + + V ++F V

50 Sbjct: 320 GSNGLGKSQLVYYIAGI--LDDKKATIKFQGIPLSAKQRLSKTSIVLQEVSLQLFAESVS 377

Query: 325 KDSN-----KARDILETINLWDDKKQFSPFQLSQGQORRLAIGSILASDSKLLLLDEPT 377  
 K+ N + +++E ++L + P LS G+Q+R+ I + L +D +L+ DEP+

55 Sbjct: 378 KEVNLGHERHPRTTEVIERLSLTLLERHPASLSGGEQQRVMIAASLLADKDILIFDEPS 437

Query: 378 YGQDAYHANMITTLTLLSYCHKNHCGVIFTSHDPHL 412  
 G D + LL+ H VI SHD L

Sbjct: 438 SGLDLLQMKALANLLMQ-LKTQHKVVILISHDEEL 471

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1439

A DNA sequence (GBSx1525) was identified in *S.agalactiae* <SEQ ID 4417> which encodes the amino acid sequence <SEQ ID 4418>. Analysis of this protein sequence reveals the following:

-1588-

Possible site: 42

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.62	Transmembrane	8 - 24 ( 1 - 30)
INTEGRAL	Likelihood = -8.17	Transmembrane	145 - 161 ( 143 - 163)
INTEGRAL	Likelihood = -6.32	Transmembrane	66 - 82 ( 62 - 84)
INTEGRAL	Likelihood = -3.77	Transmembrane	112 - 128 ( 111 - 132)
INTEGRAL	Likelihood = -2.66	Transmembrane	43 - 59 ( 43 - 59)

----- Final Results -----

bacterial membrane --- Certainty=0.5649(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAB13180 GB:Z99110 ykoE [Bacillus subtilis]

Identities = 68/177 (38%), Positives = 117/177 (65%), Gaps = 1/177 (0%)

Query: 5 LKDVLLIALLA VVLGVVYFGAGYISNAFVFPFGPIAHEVIYGIWFWAGPMALYILRKPGT 64

+K+++++++V VVY + N GPIA+E IYGIWF+ +A Y++RKPG

Sbjct: 6 VKEIVIMSVISIVFAVVYLLFTHFGNVLAGMFGPIAYEPIYGIWFIWVIAAYMIRKPGA 65

Query: 65 AIVAELLAALIEVLIGSIYGPSVLVIGITLQGLGSELGFTLFRYHNYKLPAFILSAILTSI 124

A+V+E++AAL+E L+G+ GP V+VIG +QGLG+E F R+ Y LP +L+ + +S+

Sbjct: 66 ALVSEIIAALVECLLGNPSGPMVIVIGIVQGLGAEAVFLATRWKAYSLPVLMLAGMGSSV 125

Query: 125 FSFAWSFYANGLSAFSSFSYNILMLIVRTVS-SIIFLLTKNICDQLHRSGVLNAYGI 180

SF + + +G +A+S Y ++ML++R +S +++ LL K + L +GVLN +

Sbjct: 126 ASFIYDLFVSGYAAYS PGYLLIMLVIRLISGALLAGLLGKAVSGSLAYTGVLNGMAL 182

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1440**

A DNA sequence (GBSx1526) was identified in *S.agalactiae* <SEQ ID 4419> which encodes the amino acid sequence <SEQ ID 4420>. Analysis of this protein sequence reveals the following:

Possible site: 47

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -6.69	Transmembrane	65 - 81 ( 53 - 95)
INTEGRAL	Likelihood = -6.37	Transmembrane	34 - 50 ( 31 - 54)
INTEGRAL	Likelihood = -6.10	Transmembrane	176 - 192 ( 169 - 195)
INTEGRAL	Likelihood = -3.66	Transmembrane	130 - 146 ( 130 - 151)
INTEGRAL	Likelihood = -1.97	Transmembrane	3 - 19 ( 3 - 19)
INTEGRAL	Likelihood = -0.90	Transmembrane	88 - 104 ( 88 - 104)

----- Final Results -----

bacterial membrane --- Certainty=0.3675(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

A related GBS nucleic acid sequence <SEQ ID 9757> which encodes amino acid sequence <SEQ ID 9758> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8809> and protein <SEQ ID 8810> were also identified. Analysis of this protein sequence reveals the following:

-1589-

Lipop: Possible site: -1   Crend: 8  
 McG: Discrim Score:       -4.09  
 GvH: Signal Score (-7.5): -4.38  
 Possible site: 47

>>> Seems to have no N-terminal signal sequence  
 ALOM program   count: 6 value: -6.69 threshold: 0.0  
   INTEGRAL   Likelihood = -6.69   Transmembrane   65 - 81 ( 53 - 95)  
   INTEGRAL   Likelihood = -6.37   Transmembrane   34 - 50 ( 31 - 54)  
   INTEGRAL   Likelihood = -6.10   Transmembrane   176 - 192 ( 169 - 195)  
   INTEGRAL   Likelihood = -3.66   Transmembrane   130 - 146 ( 130 - 151)  
   INTEGRAL   Likelihood = -1.97   Transmembrane   3 - 19 ( 3 - 19)  
   INTEGRAL   Likelihood = -0.90   Transmembrane   88 - 104 ( 88 - 104)  
   PERIPHERAL Likelihood = 5.30       158  
 modified ALOM score: 1.84

\*\*\* Reasoning Step: 3

----- Final Results -----

      bacterial membrane --- Certainty=0.3675(Affirmative) < succ>  
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1441

A DNA sequence (GBSx1527) was identified in *S.galactiae* <SEQ ID 4421> which encodes the amino acid sequence <SEQ ID 4422>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

      bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8811> and protein <SEQ ID 8812> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1   Crend: 2  
 McG: Discrim Score:       6.01  
 GvH: Signal Score (-7.5): 0.45  
 Possible site: 23

>>> Seems to have a cleavable N-term signal seq.

ALOM program   count: 0 value: 10.66 threshold: 0.0  
   PERIPHERAL Likelihood = 10.66       80  
 modified ALOM score: -2.63

\*\*\* Reasoning Step: 3

----- Final Results -----

      bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1590-

SEQ ID 4422 (GBS19) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 4; MW 24kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 9 (lane 6; MW 46.1kDa).

The GST-fusion protein was purified as shown in Figure 190, lane 10.

## 5 Example 1442

A DNA sequence (GBSx1528) was identified in *S.agalactiae* <SEQ ID 4423> which encodes the amino acid sequence <SEQ ID 4424>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8813> which encodes amino acid sequence <SEQ ID 8814> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length of UR: 23
      Peak Value of UR: 2.61
      Net Charge of CR: 3
McG: Discrim Score: 9.08
GvH: Signal Score (-7.5): -0.76
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 23
ALOM program count: 0 value: 5.14 threshold: 0.0
      PERIPHERAL Likelihood = 5.14 365
      modified ALOM score: -1.53

*** Reasoning Step: 3

Rule gp01

----- Final Results -----
      bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAA34476 GB:X16457 precursor polypeptide (AA -26 to 632)
      [Staphylococcus aureus]
      Identities = 93/372 (25%), Positives = 160/372 (43%), Gaps = 46/372 (12%)

Query: 9  MKKQFLKSAAILSLAVTAVSTSQPVGAIVGKDETCLRQQLGYIDSKKSGKKIDERWGEKI 68
      MKKQ + A L++A + + AIV KD +K + + K G + + + KI
Sbjct: 1  MKKQIISLGA-LAVASSLFTWDNKADAIVTKDYK--ESRVNEKSKKGATVSDYYYWKI 56

Query: 69 YNYLSYELIEANEWINRSEFQEPFYRTILSEFKDKIDSIEFYLLINLS----NIAKEDAHQ 124
      + L + A + + ++ +P Y+ ++ + YL+ + K+
Sbjct: 57 IDSLEAQFTGAIDLLENYKYGDPIYKEAKDRLMTRVLGEDQYLLKKKIDYELYKKWYKS 116

Query: 125 RNILQSLDKYEKSGIYNLDQGVYNYIYQEISSAKHKFSDGVDKIYRLDSTLFPFVSWYDK 184
      N ++ + K +YNL YN I+ + A ++F+ V +I + L F
Sbjct: 117 SNKNTNMLTFHKYNLYNLTMNEYNDIFNSLKDVAVYQFNKEVKEIEHKNVDLKKQF----- 170

Query: 185 HLDNNDNYKDNKDFKEYIALNLETTRKARLGYQIVNNHKD-GEHKDEAEI-LDILIRDIT 242
```

-1591-

```

      D  ++K KE  L++EI      Y      KD GEH E      LD+++ D
Sbjct: 171 -----DKDGEDKATKEVYDLVSEIDTLVVITYA----DKDYGEHAKELRAKLDLILGDTD 221

Query: 243 FVSKDAPGYKYIPNKRIAAKIIEDLDGIINDFFKNTGKDKP-SLEKLDTEFHKKYLNST 301
      K      I N+RI ++I+DL+ II+DFF T +++P S+ K T+ + K +
Sbjct: 222 NPHK-----ITNERIKKEMIDDLNSIIDFFMETKQNRPNISITKYDPTKHNFKKSEN 274

Query: 302 EPYSIETNLPSPNYKELKEKQIKKLEYGYK-KSSKIY--TSAHYALYSEEIDAAKELLQKV 358
      +P      N      +E K K +K+ + +K K+ K Y T      + EE      + L KV
Sbjct: 275 KP-----NFDKLVEETK-KAVKEADESWKNKTVKKYEETVTKSPVVKEEKVVEEPQLPKV 328

Query: 359 KIAKDNVNEIKS 370
      N E+K+
Sbjct: 329 ----GNQQEVKT 336

```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8814 (GBS119) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 2; MW 84.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 35 (lane 5; 2 bands).

The GBS119-GST fusion product was purified (Figure 109A; see also Figure 201, lane 6) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 109B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1443

A DNA sequence (GBSx1529) was identified in *S.agalactiae* <SEQ ID 4425> which encodes the amino acid sequence <SEQ ID 4426>. This protein is predicted to be s-adenosylmethionine synthetase (metK). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3609(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB07019 GB:AP001518 S-adenosylmethionine synthetase [Bacillus halodurans]
Identities = 266/390 (68%), Positives = 324/390 (82%), Gaps = 1/390 (0%)

Query: 4   RKLFTSESVSEGHDPKIDQISDAILDAILDQDPDAHVAETA VYTGSVHVFGEISTTAY 63
      R+LFTSESV+EGHPDKI DQISD+ILD IL++DP+A VA ET+V TG V V GEI+T+ Y
Sbjct: 7   RRLFTSESVTEGHDPKICDQISDSILDEILKEDPNARVACETSVTTGLVLVAGEITTTSTY 66

Query: 64  VDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVR-GSLEQDP 122
      VDI +VVR+TI IGY +A+YGF +E+ V S+ EQSPDIAQGVN+ALE R G +
Sbjct: 67  VDIPKVVVRTIRNIGYTRAKYGFDSSETCAVLTSIDEQSPDIAQGVNQALEARQMTDAE 126

Query: 123 LDLIGAGDQGLMFGFAVDETPELMPLPISLAHQLVKKLTDLRKSGELTYLRPDAKSQVTV 182
      ++ IGAGDQGLMFG+A +ETPELMPLPISL+H+L ++L++ RK L YLRPD K+QVTV
Sbjct: 127 IEAIGAGDQGLMFGYANNETPELMPLPISLSHKLARRLSEARKGETLPYLRPDGKTQVTV 186

Query: 183 EYDENDQPIRVDAVVISTQHDPNVINDQLHKDVIEKVINEVIPSHYLDQTKFFINPTGR 242

```

-1592-

EYDENDQ +R+D +VISTQH P VT +Q+ D+ + VI V+P +D++TK+FINPTGR  
 Sbjct: 187 EYDENDQSVRIDTIVISTQHHPVETLEQIESDLKQHVIRSVVPEELIDEETKYFINPTGR 246  
 Query: 243 FVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSAS YAARYIAKNIVAAD 302  
 FVIGGPQGD+GLTGRKIIIVDTYGGY+RHGGGAFSGKD TKVDRS +YAARY+AKNIVAA  
 Sbjct: 247 FVIGGPQGDAGLTGRKIIIVDTYGGYARHGGGAFSGKDPTKVDRSGAYAARYVAKNIVAAG 306  
 Query: 303 LAKKVEVQLAYAIGVAQPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLKRP 362  
 LA K EVQLAYAIGVA+PVS+ +DTFGTG ++EA L VR+ FDLRPAGII MLDL+RP  
 Sbjct: 307 LADKCEVQLAYAIGVAKPVISISIDTFGTGQVSEARLVELVREHFDLRPAGIIKMLDLRRP 366  
 Query: 363 IYRQTAAAYGHMGRTDIDLPERVVDKQALK 392  
 IY+QTAAAYGH GRTD++LPWE+ DK + L+  
 Sbjct: 367 IYKQTAAAYGHFGRTDVELPWEQTDKAEILR 396

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4427> which encodes the amino acid sequence <SEQ ID 4428>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3389(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 333/395 (84%), Positives = 361/395 (91%), Gaps = 1/395 (0%)

Query: 1 MSERKLFSES SVSEGHDPDKIADQISDAILDAILDQDPDAHVAEATAVYTGSVHVFGEIST 60  
 MSERKLFSES SVSEGHDPDKIADQISDAILDAIL +DP+AHVAEATAVYTGSVHVFGEIST  
 Sbjct: 1 MSERKLFSES SVSEGHDPDKIADQISDAILDAILDPEAHVAEATCVYTGSVHVFGEIST 60  
 Query: 61 TAYVDINRVVRNTIAEIGYDKAEYGFSAESVGVHPSLVEQSPDIAQGVNEALEVRGSLEQ 120  
 TAY+DINRVVR+TIAEIGY +AEYGFSAESVGVHPSLVEQS DIAQGVNEA E R +  
 Sbjct: 61 TAYIDINRVVRTIAEIGYTEAEYGFSAESVGVHPSLVEQSGDIAQGVNEAFESREG-DT 119  
 Query: 121 DPLDLIGAGDQGLMFGFAVDETPELMPLPISLAHQVLVKKLTDLRKSGELTYLRPDAKSQV 180  
 D L IGAGDQGLMFGFA++ETPELMPLPISL+HQLV++L +LRKSGE++YLRPDAKSQV  
 Sbjct: 120 DDLSHIGAGDQGLMFGFAINETPELMPLPISLSHQLVRRRLAELRKSGEISYLRPDAKSQV 179  
 Query: 181 TVEYDENDQPIRVDAVISTQHDPNVTNDQLHKDVIEKVINEVIPS HYLDDQTKFFINPT 240  
 TVEYDE+D+P+RVD VVISTQHDP TNDQ+ +DVIEKVI VIP+ YLDD TKFFINPT  
 Sbjct: 180 TVEYDEHDKPVRVDTVVISTQHDPNVTNDQIRQDVIEKVIKAVIPADYLDLDDTKFFINPT 239  
 Query: 241 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSAS YAARYIAKNIVA 300  
 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSAS YAARYIAKN+VA  
 Sbjct: 240 GRFVIGGPQGD SGLTGRKIIIVDTYGGYSRHGGGAFSGKD ATKVDRSAS YAARYIAKNLVA 299  
 Query: 301 ADLAKKVEVQLAYAIGVAQPVSVRVDTFGTGVIAEADLEAAVRQIFDLRPAGIINMLDLK 360  
 A L K EVQLAYAIGVAQPVSVRVDTFGT + EA LEAAVRQ+FDLRPAGII MLDLK  
 Sbjct: 300 AGLVTKAEVQLAYAIGVAQPVSVRVDTFGTSTVPEAVLEAAVRQVFDLRPAGIIQMLDLK 359  
 Query: 361 RPIYRQTAAAYGHMGRTDIDLPERVVDKQALKDFI 395  
 RPIY+QTAAAYGHMGRTDIDLPER++KV AL + +  
 Sbjct: 360 RPIYKQTAAAYGHMGRTDIDLPERLKNKVDALVEAV 394

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1593-

**Example 1444**

A DNA sequence (GBSx1530) was identified in *S.agalactiae* <SEQ ID 4429> which encodes the amino acid sequence <SEQ ID 4430>. This protein is predicted to be a transcriptional repressor of the biotin operon. Analysis of this protein sequence reveals the following:

```

5   Possible site: 24
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.16    Transmembrane 188 - 204 ( 188 - 204)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9755> which encodes amino acid sequence <SEQ ID 9756> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

20  >GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
    operon [Bacillus halodurans]
    Identities = 102/315 (32%), Positives = 169/315 (53%), Gaps = 18/315 (5%)

Query: 10  ILSKNNNFISGETMANQLNISRTAIWKGIKLEELGLEIESVTINKGYRLVSG-DILLPEQ 68
      +L+  ++F+SGE ++ + SRTA+WK I+ L + G E+E+V KGYR+V D + P
Sbjct: 9   LLTAGDDFVSGEKISQAIGCSRTAVWKHIEELRKSGYEVEAVQRKGYRIVKRPDQIKPHD 68

25  Query: 69  LE-----QEIGIKVSLNNSASTQLDAKMGIESKLKTPHFLAPNQKKAKGRFDRPFPTS 123
      ++   + G +++ ++ASTQ A + K H+ LA Q KGR R +++
Sbjct: 69  IQVVLETERFGREITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSKGKRMGRGWYSP 128

30  Query: 124  NQGGIYMSLLLPQNPVIEDIKPYTMVMASSAVKAI SRLTGITPEIKWVNDIYLDNKKIAG 183
      I MS++ +P +P + T++ A + V+AI TG+ +IKW ND+ +D KKI G
Sbjct: 129  PGSSISMSIIFRPQLPPQKAPQLTLLTAVAIVRAIKETTGLDSDIKWPNDLLIDGKKIVG 188

35  Query: 184  ILTEAIASVESGLVTNVIIGLGINFYIKE--FPRALTKRAGSLFTEQ-PTITRNQLITEI 240
      ILTE A +S V +VI G+GIN +E F + K A SL ++ I R LI I
Sbjct: 189  ILTEMQADQDS--VHSVIQIGIGINVNHQEEAFAEIIRKIATSLAIKKGEPIQRAPLIAAI 246

Query: 241  W---NLFFNIPLEDHLK----VYREKSLVLDRTVSFMDGQTMYSKGAIDITDKGYLVVEL 293
      LF+++ L+ ++ ++ + + + + G A ITD G L++E
40  Sbjct: 247  LKNIELFYDLYLQHGFSRIKPLWEAHASIGKRIRARMLNDVKFGVAKGITDDGVLLLED 306

Query: 294  DDGQLKTLRSGEISL 308
      DDG+L ++ S +I +
Sbjct: 307  DDGKLHSIYSADIEI 321

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4431> which encodes the amino acid sequence <SEQ ID 4432>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 34
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.49    Transmembrane 194 - 210 ( 194 - 211)

55  ----- Final Results -----
      bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:BAB05404 GB:AP001512 transcriptional repressor of the biotin
operon [Bacillus halodurans]
Identities = 98/315 (31%), Positives = 165/315 (52%), Gaps = 18/315 (5%)

```

-1594-

Query: 10 LLSQTDDFVSGEYLADQLSISRTSVWKSISLENOGIQIDSLKHKGYRMVQG-DILLPKT 68  
 LL+ DDFVSGE ++ + SRT+VWK I+ L G +++++ KGYR+V+ D + P  
 Sbjct: 9 LLTAGDDFVSGEKISQAIGCSRTAVWKHIEELRKSGYEVEAVQRKGYRIVKRPDQIKPHD 68

Query: 69 I-----SQGLGMPVTTYTPHSQSTQLDAKQIEAHNSAPRLYLAPSQEAAGRLDRQFFSA 123  
 I ++ G +TY + STQ A + + + LA Q + KGR+ R ++S  
 Sbjct: 69 IQVVLETERFGREITYLESTASTQTVALKLAQEGAKEGHIVLANEQTSKGKGRMGWYSP 128

Query: 124 STGGIYMSMYLKPNVPYADMPPTYMMVASSIVKAISRLTGIDTEIKWVNDIYLGHNHVKVAG 183  
 I MS+ +P +P P T++ A +IV+AI TG+D++IKW ND+ + K+ G  
 Sbjct: 129 PGSSISMSIIFRPQLPPQKAPQLTLLTAVAIVRAIKETTGLDSIDKWPNDLLIDGKKIVG 188

Query: 184 ILTEAITSVETGLITDVIIGVGLNFFVTD--FPEAIAQKAGSLFTEK-PTITRNDLIIDI 240  
 ILTE + + VI G+G+N + F E I + A SL +K I R LI I  
 Sbjct: 189 ILTE--MQADQDSVHSVIQIGGINVNHQEEAFEEIRKIATSLAIKKGEPIQRAPLIAAI 246

Query: 241 WK-----LFLSIPVKDHVKVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHILIVQF 293  
 K L+L +++ ++ + K++ + K +A +TD G L+++  
 Sbjct: 247 LKNIELFYDLYLQHGFSRIKPLWEAHAISIGKRIRARMLNDVKFGVAKGITDDGVLLLED 306

Query: 294 ENGDQLTLRSGEISL 308  
 ++G L ++ S +I +  
 Sbjct: 307 DDGKLHSIYSADIEI 321

An alignment of the GAS and GBS proteins is shown below.

Identities = 191/311 (61%), Positives = 257/311 (82%)

Query: 1 MKTYEKIYQILSKNNNFISGETMANQNLNISRTAIWKGIKTLEELGLEIESVTNKGYRLVS 60  
 MKT EKIIYQ+LS+ ++F+SGE +A+QL+ISRT++WK IK+LE G++I+S+ +KGYR+V  
 Sbjct: 1 MKTSEKIYQLLSQTDDFVSGEYLADQLSISRTSVWKSISLENOGIQIDSLKHKGYRMVQ 60

Query: 61 GDILLPEQLEQEIGIKVSLNNSASTQLDAKMGIESKLKTPHLFLAPNQKKAKGRFDRPF 120  
 GDILLP+ + Q +G+ V+ +S STQLDAK GIE+ P L+LAP+Q+ AKGR DR F  
 Sbjct: 61 GDILLPKTISQGLGMPVTTYTPHSQSTQLDAKQIEAHNSAPRLYLAPSQEAAGRLDRQF 120

Query: 121 FTSNQGGIYMSLLLQPNVPIDIKPYTMVASSAVKAISRLTGITPEIKWVNDIYLDNKK 180  
 F+++ GGIYMS+ L+PNVP D+ PYT+MVASS VKAISRLTGI EIKWVNDIYL N K  
 Sbjct: 121 FSASTGGIYMSMYLKPNVPYADMPPTYMMVASSIVKAISRLTGIDTEIKWVNDIYLGHNH 180

Query: 181 IAGILTEAIASVESGLVTNVIIGLGINFYIKEFPRALTKRAGSLFTEQPTITRNQLITEI 240  
 +AGILTEAI SVE+GL+T+VIIG+G+NF++ +FP A+ ++AGSLFTE+PTITRN LI +I  
 Sbjct: 181 VAGILTEAITSVETGLITDVIIGVGLNFFVTDFFPEAIAQKAGSLFTEKPTITRNDLIIDI 240

Query: 241 WNLFFNIPLDHLKVYREKSLVLDRTVSFMDGQTMYSKKAIDITDKGYLVVELDDGQLKT 300  
 W LF +IP++DH+KVY+EKSLVL++ V+F++ AID+TD+G+L+V+ ++G L+T  
 Sbjct: 241 WKLFSLIPVKDHVKVYKEKSLVLNKQVTFIENSQEKRAIAIDLTDQGHILIVQFENGDLQ 300

Query: 301 LRSGEISLSSW 311  
 LRSGEISLSSW  
 Sbjct: 301 LRSGEISLSSW 311

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1445

A DNA sequence (GBSx1531) was identified in *S.agalactiae* <SEQ ID 4433> which encodes the amino acid sequence <SEQ ID 4434>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -2.76 Transmembrane 3 - 19 ( 3 - 20)

----- Final Results -----

-1595-

```

bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1446

10 A DNA sequence (GBSx1532) was identified in *S.agalactiae* <SEQ ID 4435> which encodes the amino acid sequence <SEQ ID 4436>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -2.28    Transmembrane    24 - 40 ( 24 - 40)
15
----- Final Results -----
        bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4437> which encodes the amino acid sequence <SEQ ID 4438>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
25 >>> Seems to have a cleavable N-term signal seq.
    INTEGRAL    Likelihood = -1.91    Transmembrane    58 - 74 ( 58 - 75)
----- Final Results -----
30        bacterial membrane --- Certainty=0.1765(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 37/67 (55%), Positives = 54/67 (80%), Gaps = 3/67 (4%)
35
Query: 1  MTKRQFIFMALLCSFETYFFNQSVMDGSWIFAFWGVLLLRDLQKVYAISKFTKELIK-- 58
          MT RQF+FMA +C+FETYFFN ++ G+++FA+FWG+LL RDL++V+ I++ TK ++K
Sbjct: 36  MTIRQFLFMAFVCAFETYFFNDLLLSGNLFEALFWGLLLFRDLRRVHTINQLTKTILKTA 95
40
Query: 59  -STKKKD 64
          S KKKD
Sbjct: 96  NSPKKKD 102

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1447

A DNA sequence (GBSx1533) was identified in *S.agalactiae* <SEQ ID 4439> which encodes the amino acid sequence <SEQ ID 4440>. This protein is predicted to be DNA polymerase III, gamma subunit (dnaZX). Analysis of this protein sequence reveals the following:

```

50 Possible site: 60
    >>> Seems to have no N-terminal signal sequence

```

-1596-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1567(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4441> which encodes the amino acid sequence <SEQ ID 4442>. Analysis of this protein sequence reveals the following:

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.59 Transmembrane 232 - 248 ( 232 - 249)

## ----- Final Results -----

bacterial membrane --- Certainty=0.1235(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 408/558 (73%), Positives = 473/558 (84%), Gaps = 6/558 (1%)

Query: 1 MYQALYRKYSQTFDEMVGQSVISTTLKQAVSSKKISHAYLFSGPRGTGKTSAAKIFAKA 60  
 MYQALYRKYSQTFDEMVGQSVISTTLKQAV S KISHAYLFSGPRGTGKTSAAKIFAKA  
 Sbjct: 1 MYQALYRKYSQTFDEMVGQSVISTTLKQAVESGKISHAYLFSGPRGTGKTSAAKIFAKA 60

Query: 61 MNCNPQINGEPCNHCDICRDTNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK 120  
 MNCNPQ++GEPCN CDI<sup>~</sup>CRDTNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK  
 Sbjct: 61 MNCNPQVDGEPCNQCDICRDTNGSLEDVIEIDAASNNGVDEIRDIRDKSTYAPSRATYK 120

Query: 121 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIKLL 180  
 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIK  
 Sbjct: 121 VYIIDEVHMLSTGAFNALLKTLLEPTENVVFILATTELHKIPATILSRVQRFEFKAIKQK 180

Query: 181 AIRDHLAQILDKEAISYDL<sup>~</sup>DALTLVARRAEGGMRDALSILDQALS<sup>~</sup>LAKDNHISLDVAEEI 240  
 AIR+HLA +LDKE I+Y++DAL L+ARRAEGGMRDALSILDQALS+ DN +++ +AEEI  
 Sbjct: 181 AIREHLAWVLDKEGIAYEVDALNLIARRAEGGMRDALSILDQALS<sup>~</sup>LPDNQVAIAIAEEI 240

Query: 241 TGSISLSAIDYVSNI<sup>~</sup>LADHTTEALAKLEIVFDSGKSMSRFATDLLMYRLDLLVVQAGGE 300  
 TGSIS+ A+ DYV + T+ALA LE I+DSGKSMSRFATDLL YLRDLLVV+AGG+  
 Sbjct: 241 TGSISILALGDYVRYVSQE<sup>~</sup>QATQALAALETIYDSGKSMSRFATDLLTYLRDLLVVKAGGD 300

Query: 301 DSHSSDTFIANLN<sup>~</sup>VKQDILFEMIDKVTSLVPEIKNGSHPKVYAEMMTIQLSEMVEKNSS- 359  
 + S F NL++ D +F+MI VTS LPEIK G+HP++YAEMMTIQL++ + S  
 Sbjct: 301 NQRQSAVFDTNLSLSIDRIFQMITVTVSHLPEIKKGTHPRIYAEMMTIQLAQKEQILSQV 360

Query: 360 NIPADVTAELDSLRLRELKSLKNEMSQL-SRADQSSSTQKVKNNTFTFKVDRTKILTIM 418  
 N+ ++ +E+++L+ EL LK ++SQL SR D + + K K KT +++VDR IL IM  
 Sbjct: 361 NLSGELISEIETLKNELAQ<sup>~</sup>LKQQLSQLQSRPDSLARS<sup>~</sup>DKTK--PKTTSYRVDRVTILKIM 418

Query: 419 EETVVD<sup>~</sup>SQRSREYLEALKSAWNEILDNITAQDRALLMGSEPVLANSENAILAFDAAFNAE 478  
 EETV +SQ+SR+YL+ALK+AWNEILDNI+AQDRALLMGSEPVLANSENAILAF+AAFNAE  
 Sbjct: 419 EETVRNSQQSRQYLDALKN<sup>~</sup>AWNEILDNISAQDRALLMGSEPVLANSENAILAF<sup>~</sup>EEAFNAE 478

Query: 479 QAMKRTDLNDIFGNIMSKAAGFSPN<sup>~</sup>LAVPRNDFNQIRSDFAKKMKAQK--TETEPEVNH 536  
 Q M R +LND+FGNIMSKAAGFSPN<sup>~</sup>LAVPR DF IR +FA++MK+QK + E EV  
 Sbjct: 479 QVMSRNNLNDMFGNIMSKAAGFSPN<sup>~</sup>LAVPRITDFQHIRKEFAQMKSQKDSVQEEQEVAL 538

Query: 537 QIPEDFSYLAERIAIVED 554  
 IPE F +L ++I ++D  
 Sbjct: 539 DIPEGDFLLDKINTIDD 556

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1597-

**Example 1448**

A DNA sequence (GBSx1534) was identified in *S.agalactiae* <SEQ ID 4443> which encodes the amino acid sequence <SEQ ID 4444>. Analysis of this protein sequence reveals the following:

```

Possible site: 40
5  >>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----
                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
10                bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06927 GB:AP001518 unknown conserved protein [Bacillus halodurans]
Identities = 67/143 (46%), Positives = 96/143 (66%)
15
Query: 8  ENYQLLLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFDFGEELILGPFQGGV 67
          E Y L+ Q AL E++A+ANL+NASA+L L + GFYL EL+LGPFQG
Sbjct: 13 EKYSLVTKQLAALLEGESDAIANLANASALLYHFLEEVNWWGFFYLKEGELVLGPFQGLP 72

20
Query: 68 SCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDKAMSEIVVPMFKNGKLLGVLDL 127
          +CV I +G+GVCG +A+ +T+ V+DV + +I+CD+ + SEIV+P+F+NG L GVLD+
Sbjct: 73 ACVRPIPIGRGVCCTAAKEEQTVRVEDVHQFPGHIACDAASRSEIVIPLFQNGVLYGVLDI 132

25
Query: 128 DSSLVADYDEIDQEYLEKFVGIL 150
          DS + + E +Q LE FV +L
Sbjct: 133 DSPSLNRFSEEEQALLESFVDVL 155

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4445> which encodes the amino acid sequence <SEQ ID 4446>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 27
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.1753 (Affirmative) < succ>
35                bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
                bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 122/164 (74%), Positives = 144/164 (87%)
40
Query: 1  MNKSKKIENYQLLLLLQAQALFSDETNALANLSNASAMLNAMLPNSVFTGFYLFDFGEELIL 60
          MNKSKKIE YQL++ QA+ LF++E+NALANLSNASA+LN LPNSVFTGFYLFDFG+ELIL
Sbjct: 1  MNKSKKIEQYQLMIAQAKELFANESNALANLSNASALLNMTLPNSVFTGFYLFDFGQELIL 60

45
Query: 61 GPFQGGVSCVHITLGKGVCGESAQTAKTLIVDDVTKHANYISCDKAMSEIVVPMFKNGK 120
          GPFQG VSCVHI LGKGVCGESAQ+ +T+I++DV +HANYISCD+ AMSEIVVPM K G
Sbjct: 61 GPFQGRVSCVHIKLGKGVCGESAQSRRTIIINDVKQHANYISCDAAAMSEIVVPMVKEGH 120

50
Query: 121 LLGVLDLDSSLVADYDEIDQEYLEKFVGILVEHTIWNLDMFGVE 164
          L+GVLDLDSSLVADYDE+DQEYLE FV + +E T + +MFGV+
Sbjct: 121 LIGVLDLDSSLVADYDEVDQEYLEAFVDLFLEKTTFTFNMFVGK 164

```

SEQ ID 4444 (GBS282) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 9; MW 19.8kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 6; MW 44.8kDa) and in Figure 63 (lane 7; MW 47kDa).

-1598-

The GBS282-GST fusion product was purified (Figure 211, lane 4; see also Figure 225, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 269), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1449

A DNA sequence (GBSx1535) was identified in *S.galactiae* <SEQ ID 4447> which encodes the amino acid sequence <SEQ ID 4448>. This protein is predicted to be uridine kinase (udk). Analysis of this protein sequence reveals the following:

```

10   Possible site: 24
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20   >GP:CAB14675 GB:Z99117 uridine kinase [Bacillus subtilis]
    Identities = 133/207 (64%), Positives = 167/207 (80%)

    Query: 1   MRKKPIIIGVTGGSGGKTSVSRILSNFPDQKITMIEHDSYYKDQSHLTFEERVKTNYD 60
              M K P++IG+ GGSG GKTSV+R+I   F   I MI+ D YYKDQSHL FEER+ TNYD
    Sbjct: 1   MGKNPVVIGIAGGSGSGKTSVTRSIYEQFKGHSILMIQDLYYKDQSHLPFEERLNTNYD 60

25   Query: 61   HPLAFDTNLMIEQLNELIEGRPVDIPVYDYTKHTRSDRTIRQEPQDVIIIEGILVLEDQR 120
              HPLAFD + +IE + +L+ RP++ P+YDY HTRS+ T+ EP+DVII+EGILVLED+R
    Sbjct: 61   HPLAFDNDYLIEHIQDLNRYPIEKPIYDYKLHTRSEETVHVEPKDVIILEGILVLEDKR 120

30   Query: 121  LRDLMDIKLFVDTDDDIRIIRRIKRDMEERDRSLDSIEQYTEVVKPMYHQFIEPTKRYA 180
              LRDLMDIKL+VDTD D+RIIRRI RD+ ER RS+DS+IEQY VV+PM++QF+EPTKRYA
    Sbjct: 121  LRDLMDIKLYVDTDADLRRIIRIMRDINERGRSIDSIEQYVSVVRPMHNQFVEPTKRYA 180

    Query: 181  DIVIPEGVSNIVAIDLINTKVASILNE 207
              DI+IPEG N VAIDL+ TK+ +IL +
35   Sbjct: 181  DIIIEGGQNHVAIDLMVTIKITILEQ 207

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4449> which encodes the amino acid sequence <SEQ ID 4450>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 39
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9151> which encodes the amino acid sequence <SEQ ID 9152>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 35
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
55   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1599-

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/207 (83%), Positives = 193/207 (92%)

5  
 Query: 1 MRKKPIIIGVTTGGSGGKTSVSRAILSNNFPDQKITMIEHDSYYKDQSHLTFEERVKTNYD 60  
 M KKPIIIGVTTGGSGGKTSVSRAIL +FP+ +I MI+HDSYYKDQSH++FEERVKTNYD  
 Sbjct: 5 MLKKPIIIGVTTGGSGGKTSVSRAILDSFPNARIAMIQHDSYYKDQSHMSFEERVKTNYD 64

10  
 Query: 61 HPLAFDTNLMIEQLNELIEGRPVDPVYDYTKHTRSDRTIRQEPQDVIIIVEGILVLEDQR 120  
 HPLAFDT+ MI+QL EL+ GRPVDIP+YDY KHTRS+ T RQ+PQDVIIIVEGILVLED+R  
 Sbjct: 65 HPLAFDTDFMIQQLKELLAGRVPDIPIYDYKKHTRSNITFRQDPQDVIIIVEGILVLEDER 124

15  
 Query: 121 LRDLMIDIKLFVDITDDDIRIIRRIKRDMEERDRSLDSIIQYTEVVVKPMYHQFIEPTKRYA 180  
 LRDLMIDIKLFVDITDDDIRIIRRIKRD ER RSL+SII+QYT VVKPMYHQFIEP+KRYA  
 Sbjct: 125 LRDLMIDIKLFVDITDDDIRIIRRIKRDMMERGRSLESIIDQYTSVVKPMYHQFIEPSKRYA 184

20  
 Query: 181 DIVIPEGVSNIVAIDLINTKVASILNE 207  
 DIVIPEGVSN+VAID+IN+K+ASIL E  
 Sbjct: 185 DIVIPEGVSNVVAIDVINSKIASILGE 211

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1450**

25 A DNA sequence (GBSx1536) was identified in *S.agalactiae* <SEQ ID 4451> which encodes the amino acid sequence <SEQ ID 4452>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5083(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12572 GB:Z99108 similar to RNA helicase [Bacillus subtilis]  
 Identities = 140/343 (40%), Positives = 202/343 (58%), Gaps = 9/343 (2%)

40  
 Query: 10 QDKLTQRQFDDLTDIQNKLFQPTDGDNIIGISPTGTGKTLAYLFPTLLKLQPK-KSQQ 68  
 Q+ F T +Q + Q I DG +++ SPTGTGKTLAY P L +++P+ K Q  
 Sbjct: 16 QENWNASGFQKPTFPVQEQAQLIMDGKDVIASPTGTGKTLAYALPVLERIKPEQKHQA 75

45  
 Query: 69 LILAPNSELAQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGPEILIGTAGRVFE 128  
 +ILAP+ EL QIF V ++W L A + G++ K+Q+E+LKK P I++GT GRVFE  
 Sbjct: 76 VILAPSRELVMQIFQVIQDWKAGSELRAASLIGGANVKKQVEKLKKHPIIVGTPGRVFE 135

50  
 Query: 129 LVKLKKIKMMNINTIVLDEFDELGLDGSQYHFVDNIINRVPRDQMIYISATNKLDNS--- 185  
 L+K KK+KM + TIVLDE D+L+ + II RD+Q++ SAT K +  
 Sbjct: 136 LIKAKKLMHEVKTIVLDETDQLVLPEHRETMQIIKTTLRDRQLLCSATLKKETEDVL 195

55  
 Query: 186 -KLADNTTITIDLSNQKLDT--IKHYITVDKRETDLLRKFSNIPDFRGLVFFNSLSDLG 242  
 +LA + + K + +KH Y+ D+R++ LL+K S + + LVF + +L  
 Sbjct: 196 RELAQEPEVLKVQSKAEAGKVKHQYLI CDQDKVKLLQKLSRLEGMQALVFVRDIGNLS 255

60  
 Query: 243 ACEERLQFNRAVSLASDINIKPRKVILEKFKNHDISLLGLTDLVARGIDIDNLEYVIN 302  
 E+L ++ L S+ R I+ F++ + LLL TD+ ARG+DI+NL YVI+  
 Sbjct: 256 VYAEKLAYHHVELGVLHSEAKMERAKI IATFEDGEFPLLLATDIAARGLDIENLPYVIH 315

60  
 Query: 303 FDIARDKETYYTHRSRGTGRMGKEGCVITFVTHKEELKQLKKYA 345  
 DI D++ Y HRSRGTGR GKEG V++ VT EE K LKK A  
 Sbjct: 316 ADIP-DEGIVHRSRGTGRAGKEGNVLSLVTLEESK-LKKMA 356

-1600-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4453> which encodes the amino acid sequence <SEQ ID 4454>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3847(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 273/358 (76%), Positives = 312/358 (86%)

```

Query: 1  MITKFPDQWQDKLTQRQFDDLTDIQNKLFQPIITDGDNI LGISPTGTGKTLAYLFPTLLKL 60
          MITKFP QWQ+KL Q F LT IQ + FQPI DG N LGISPTGTGKTLAY+FP LL L
Sbjct: 12 MITKFPDQWQEKLDQVAFTHLTPIQEQAQPIVDGKNFLGISPTGTGKTLAYVFPNLLAL 71

Query: 61  QPKKSQQLLILAPNSELAQIFDVTKEWAEPLGLTAQLFLSGSSQKRQIERLKKGP EILI 120
          PKKSQQLLILAPN+ELAQIF+VTK+WA+PLGLTAQLF+SG+SQKRQIERLKKGP EILI
Sbjct: 72  TPKKSQQLLILAPNTELAQIFEVTKDWAQPLGLTAQLFISGTSQKRQIERLKKGP EILI 131

Query: 121  GTAGRVFELVKLKKIKMMNINTIVLDEFDELLGDSQYHFVDNIINRVPRDQOMIYISATN 180
          GT GR+FEL+KLKKIKMM++NTIVLDE+DELLGDSQY FV I + VPRD QM+Y+SATN
Sbjct: 132  GTPGRIFELIKLKKIKMMSVNTIVLDEYDELLGDSQYDFVQKISHYVPRDHQMVMYSATN 191

Query: 181  KLDNSKLADNTITIDLSNQKLDTIKHYIITVDKRERTDLLRKFSNIPDFRGLVFFNSLSLD 240
          K+D + LA NT IDLS Q D I+H+Y+ VDKRERTDLLRK F+NIP FR LVFFNSLSLD
Sbjct: 192  KVDQTS LAPNTFCIDLSEQTNDIAIQHYFLMVDKRERTDLLRKFTNIPHFRLVFFNSLSLD 251

Query: 241  LGACEERLQFNRAVSLASDINIKFRKVILEKFKNHDISLLLGTDLVARGIDIDNLEYV 300
          LGA EERLQ+N A+AVSLASDIN+KFRK ILEKFK+H +SLLL TDLVARGIDIDNL+YV
Sbjct: 252  LGATEERLQYNGAAVSLASDINVKFRKTILEKFKSHQLSLLLATDLVARGIDIDNLDYV 311

Query: 301  INFDIARDKETYTHRSGRGTGRMGKEGCVITFVTHKEELKQLKKYATVTELVLHNQKLH 358
          I+FD+ARDKE YTHR+GRTGRMGK G VITFV+H E+LK+LKK+A V+E+ L NQ+LH
Sbjct: 312  IHFDVARDKENYTHRAGRTGRMGKSGIVITFVSHPEDLKKLKKFAKVSEISLKNQQLH 369

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1451

A DNA sequence (GBSx1537) was identified in *S.agalactiae* <SEQ ID 4455> which encodes the amino acid sequence <SEQ ID 4456>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -1.38 Transmembrane 15 - 31 ( 13 - 31)

----- Final Results -----

bacterial membrane --- Certainty=0.1553(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1601-

**Example 1452**

A DNA sequence (GBSx1538) was identified in *S.agalactiae* <SEQ ID 4457> which encodes the amino acid sequence <SEQ ID 4458>. This protein is predicted to be peptidoglycan GlcNAc deacetylase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -8.92    Transmembrane    4 - 20 ( 1 - 26)

      ----- Final Results -----
10          bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
      [Streptococcus pneumoniae]
      Identities = 133/431 (30%), Positives = 228/431 (52%), Gaps = 20/431 (4%)

20   Query: 5   IIGIFSLIIIAILAWQGFSLKHK--BIKLQQA VVEKEIRIAEKTVEVVKRQKTERVLF 62
      +IGI ++ I + + F + K E K++ EK+ +++E + RQ V+
      Sbjct: 21 LIGILAISICLLGGFIAFKYIQQKSFEQKIESLKKEKDDQLSEGNQKEHFRQGOAEVIAY 80

      Query: 63 EPKGYDKSLADILKWNQKSFHKKFYDNQYIILRPQLADSNFANVKKLSIYQILYQKEK 122
      P +K +S+ NQ + + DN Q +S V ++ + +Y
25   Sbjct: 81 YPLQGEKVISSVRELINQDVKDLESKDNLVFYYTEQ-EESGLKGVVNRNVTKQIYDLVA 139

      Query: 123 GSMFQKSSRLRLRTYLLDQNKPFELDELLAHNISGFKAILENIAPGTQLK--EHDSNKEF 180
      + + L L ++ +PF LD+L + + +++ + + K E D +++
30   Sbjct: 140 FKIEETKTSLSGKVHLTEDGQPFTLDQLFSDASKAKEQLIKELTSFIEDKKIEQDQSEI 199

      Query: 181 LKTGRVTD----GLDVKGKLI-----NDLKLPLDKLYNVIDESYLKSSDLDLVS 227
      +K D D KD ++I+ ++ LP+ ++VI SYL D L
      Sbjct: 200 VKNFSDQDLNFWNFYKDSQIILYPSFVVENLEELALPVSAFFDVIQSSYLLEKDAALYQ 259

35   Query: 228 NLKAKAPR--VALTFDDGPNKTTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAE 285
      + K + VALTFDDGPN TTP+ LE L +Y+ KATFFV+G++ G+ D+++R+ +E
      Sbjct: 260 SYFDKKHKVVALTFDDGPNPATTQVLETLAKYDIKATFFVLGKNVSGNEDLVKRIKSE 319

40   Query: 286 GHEIGNHTWDHPNLTCLPAEKIKEEIHKTNDLIMKATGQKFPVYLRPPYGATNATVKTVTG 345
      GH +GNH+W HP L++L ++ K++I T D++ K G +RPPYGA ++
      Sbjct: 320 GHVVGNSWSHPILSQLSLDEAKKQITDTEVDLTKVLGSSSKLMRPPYGAITDDIRNSLD 379

      Query: 346 LKEMLWSVDTEDWKNHNTQAMMTNIKKQLRPGGVILMHDIHQTTIDALPTIMDYLTQGY 405
      L ++W VD+ DWK+ N +++T I+ Q+ G ++LMHDIH T++ALP +++YL QGY
45   Sbjct: 380 LSFIMWDVDSLWKSNEASILTEIQHQVANGSIVLMHDIHSPTVNALPRVIEYLNQGY 439

      Query: 406 YFVTVGELYST 416
      FVT+ E+ +T
      Sbjct: 440 TFVTIPEMLNT 450
50

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4459> which encodes the amino acid sequence <SEQ ID 4460>. Analysis of this protein sequence reveals the following:

```

55   Possible site: 22
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -12.58    Transmembrane    6 - 22 ( 1 - 27)

      ----- Final Results -----
60          bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

-1602-

The protein has homology with the following sequences in the databases:

```

!GB:AJ251472 peptidoglycan GlcNAc deacetylase [Strep...    239  4e-62

>GP:CAB96552 GB:AJ251472 peptidoglycan GlcNAc deacetylase
  [Streptococcus pneumoniae]
  Identities = 136/438 (31%), Positives = 230/438 (52%), Gaps = 23/438 (5%)

Query: 3  KLNVLVGLLSILMLSLAI---VFINRWKLNEDSQRIVLAEKKKNTSDLVKAVKHKK 58
          K  +L+ L+ IL +S+ +   +   ++   Q+I  +K+K+   +H ++
Sbjct: 13  KTRHVLALIGILAIISICLLGGFIAFKIYQQKSFEQKIESLKKKDDQLSEGNQKEHFRQ 72

Query: 59  DQKDYFFSPK--QADDFVDNLPVSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISK 116
          Q  +  + P++ +   + +   + K  S  L+  +  + S L+ V  ++K
Sbjct: 73  GQAEVIAYYPLQGEKVISSVRELINQDVKDKLESKDNLVFYYTEQEEGLKGVVNRNVTK 132

Query: 117 IVYQKKFFHLAKKSEKVISTYHVTDDLKPFQVKDLVSGHL---ERIQEEVEKKYPDAGFN 173
          +Y  F  + + +  +  H+T+D +PF +  L  S      E++ +E+  D
Sbjct: 133 QIYDLVAFKIEETKTSGLKVHLTEDGQPFLLDQLFSDASKAKEQLIKELTSFIEDKKIE 192

Query: 174 SDKYNGLKESNS---LLSDGFVEKSGNLIFD-----KKLTIPLTTLFDVINPDFLAN 222
          D+  + ++ S  L +  F+ K  +I      +++ +P++  FDVI  +L
Sbjct: 193 QDQSEQIVKNFSDQDLSAWNFDYKDSQIILYPSPVVENLEBIALPVSAFFDVIQSSYLLE 252

Query: 223 SDRAAYDNYRTYKEQHPKKLVALTFDDGPDPTTTPQVLDILAKYQAKGTFMIGSKVNN 282
          D A Y +Y  K  Q  K+VALTFDDGP+P TTPQVL+ LAKY  K TFF++G  V  N
Sbjct: 253 KDAALYQSYFDKKHQ---KVVALTFDDGPNPATTPQVLETLAKYDIKATFFVLGKNVSGN 309

Query: 283 ENLTKRVSDAGHEIANHTWDHPNLTNLSVSEIQHQVNMTNQAIEKACGKKPRYLPPYGA 342
          E+L KR+  GH + NH+W HP L+ LS+ E + Q+  T  + K  G  + +RPPYGA
Sbjct: 310 EDLVKRIKSEGHVVGNSWSHPILSQLSLDEAKKQITDTEVDLTKVLGSSSKLMRPPYGA 369

Query: 343 TNATVQQSSGLTQMLWTVDTRDWHNSTDGTMTNVKNQLQPGGVLMHDIHQTTINALPT 402
          ++ S  L+ ++W VD+ DW++ +  I+T +++Q+  G +VLMHDIH  T+NALP
Sbjct: 370 ITDDIRNSLDLSFIMWDVDSLDWKSNEASILTEIQHQVANGSIVLMHDIHSPTVNALPR 429

Query: 403 VMEYLKAEGYECVTVSEL 420
          V+EYLK +GY  VT+ E+
Sbjct: 430 VIEYLKNQGYTFVTIPEM 447

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 169/420 (40%), Positives = 259/420 (61%), Gaps = 12/420 (2%)

Query: 4  LIIGIFSLIIAILAWQGFSLKHKEIKLQQA VVEKEIRIAEKTVEVVKRQKTER--VLF 61
          +++G+ S++++ LA  + K  E  +  + EK+  ++ ++ VK  K ++  +
Sbjct: 7  ILVGLLSILMLS-LAIVFINRWKLNEDSQRIVLAEKKKNTSDLVKAVKHKKDQKDYFF 65

Query: 62  LEPKGYDKSLSDILKWNQKSFEHKKFYDNQYIILRPQLADSNFANVKLSIYQILYQKE 121
          P          D L      S  KK  D + I++RP+L  S+  +V  L+I +I+YQK+
Sbjct: 66  FSPIKQADDFVVDNLP---VSLYKKKNSDKELILVRPKLQSSHLRSVNTLTISKIVYQKK 122

Query: 122 KGSMFOKSSRLRLRTYLLDQNKKPFELDELLAHNISGFKAILIENIAPGTQLKEHDSNKEFL 181
          + +KS +++ TY +  + KPF++ +L++ ++  + +E  P          N
Sbjct: 123 FFHLAKKSEKVISTYHVTDDLKPFQVKDLVSGHLERIQQEEVEKKYPDAGFNSDKYNGLKE 182

Query: 182 KTGRVTDGLDVKGKLIIND-LKLPLDKLYNVIDESYLKSSDLDLVSNL---KAKAPR-- 235
          ++DG +VK G LI +  L +PL  L++Vi+  +L +SD      N  K + P+
Sbjct: 183 SNSLLSDGFVEKSGNLIFDKKLTIPLTTLFDVINPDFLANSRAAYDNYRTYKEQHPKKL 242

Query: 236 VALTFDDGPNKETTTPKALEILKRYNAKATFFVMGQSAVGHTDILQRMHAEIGHEIGNHTWD 295
          VALTFDDGP+  TTP+ L+IL +Y AK TFF++G  V + ++ +R+  GHEI NHTWD
Sbjct: 243 VALTFDDGPDPTTTPQVLDILAKYQAKGTFMIGSKVNNENLTKRVSDAGHEIANHTWD 302

Query: 296 HPNLTKLPAEKIKEEIHKTNDLIMKATGQKPVYLRPPYGATNATVKTVTGLKEMLSVDT 355
          HPNLT L  +I+ +++ TN  I KA G+KP YLRPPYGATNATV+  +GL +MLW+VDT
Sbjct: 303 HPNLTNLSVSEIQHQVNMTNQAIEKACGKKPRYLPPYGATNATVQQSSGLTQMLWTVD 362

```

-1603-

Query: 356 EDWKNHNTQAMMTNIKKQLRPGGVILMHDIHQTTIDALPTIMDYLTQGYFVTVGELY 415  
 DW+NH+T +MTN+K QL+PGGV+LMHDIHQTTI+ALPT+M+YL +GY VTV ELY+  
 Sbjct: 363 RDWENHSTDGIMTNVKNQLQPGGVILMHDIHQTTINALPTVMEYLKAEGYECVTVSELYA 422

5 GBS281d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 8-10; MW 71.5kDa) and in Figure 187 (lane 10; MW 71kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 152 (lane 12; MW 46.5kDa) and in Figure 183 (lane 2; MW 46kDa). Purified GBS281d-GST is shown in lane 6 of Figure 237.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1453

A DNA sequence (GBSx1539) was identified in *S.agalactiae* <SEQ ID 4461> which encodes the amino acid sequence <SEQ ID 4462>. Analysis of this protein sequence reveals the following:

15 Possible site: 57  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 20 bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4463> which encodes the amino acid sequence <SEQ ID 4464>. Analysis of this protein sequence reveals the following:

25 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.2799(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 311/475 (65%), Positives = 389/475 (81%)

35 Query: 1 MTKEYQNYVNGEWKSSVNQIEILSPIDSSLGFPVAMPTRREEVDHAMKAGREALPAWAALT 60  
 + K+Y+N VNGEWK S N+I I +P LG VPAMT+ EVD + ++AL W AL+  
 Sbjct: 1 LAKQYKNLVNGEWKLSENEITTYAPATGEEELGSPVAMPQAEDAVYASAKKALSDWRALS 60

40 Query: 61 VYERAQYLHKAADIIERDKKEIATVLAKEISKAYNASVTEVVRTADLIRYAAEEGIRLST 120  
 ERA YLHKAADI+ RD E+I +L+KE++K + A+V+EV+RTA++I YAAEEG+R+  
 Sbjct: 61 YVERAAYLHKAADILVRDAEKIGAILSKEVAKGHKAHVSEVIRTAETIINYAAEEGLRMEG 120

45 Query: 121 SADEGGKMDASTGHKLAVIRRPQVGVILAIAPYNYPNVNLGSKIAPALIGGNVVMFKPPT 180  
 EGG +A++ K+A++RR+PVG+VLAI+P+NYPVNL+GSKIAPALI GNVV KPPT  
 Sbjct: 121 EVLEGGSFEEAASKKIAIVRREPVLGLVLAISPFPNYPNLAGSKIAPALIAGNVVALKPPT 180

Query: 181 QGSVSGVLAKAFAEAGLPAGVFNTITGRGSEIGDYIVEHEEVNFINFSTGTPVGKRIGK 240  
 QGS+SGL+LA+AFAEAG+PAGVFNTITGRGS IGDYIVEHE V+FINFTGSTP+G+ IGK  
 50 Sbjct: 181 QGSISGLLLAEFAEAGIPAGVFNTITGRGSEIGDYIVEHEAVSFINFSTGSTPIGEGIGK 240

Query: 241 LAGMRPIMLELGGKDAGVVLADADLDNAKQIVAGAYDYSQRCTAIKRVLVVEEVADEL 300  
 LAGMRPIMLELGGKD+ +VL DADL AAK IVAGA+ YSGQRCTA+KRVLV+++VAD+L  
 Sbjct: 241 LAGMRPIMLELGGKDSAIVLEADLALAAKNIVAGAFGYSGQRCTAVKRVLVMDKVADQL 300

55 Query: 301 AEKISENVAKLSVGDPDFDNATVTPVIDDNSADFIESLVVDARQKGAKELNEFKRDGRLLT 360

-1604-

A +I V KLSVG P D+A +TP+ID ++ADF+E L+ DA KGA L F R+G L++  
 Sbjct: 301 AAEIKTLVEKLSVGMPEDDADITPLIDTSAADFVEGLIKDATDKGATALTAFNREGNLIS 360

Query: 361 PGLFDHVTIDMKLAWEEPFGPILPIIRVKDAEEAVAIAANKSDFGLQSSVFTTRDFQKAFDI 420  
 P LFDHVT DM+LAWEEPFGP+LPIIRV EEA+ I+N+S++GLQ+S+FT +F KAF I  
 Sbjct: 361 PVLFDHVTIDMRLAWEEPFGPVLPIIRVTTVEEAIKISNESEYGLQASIFTTNPFAFGI 420

Query: 421 ANKLEVGTVHINNKTGRGPDNFPFLGLKSGAGVQGIYSIEAMTNVKSIVFDMK 475  
 A +LEVGT VH+NNKT RG DNFPFLG K SGAGVQG++YSIEAMT VKS+VFD++  
 Sbjct: 421 AEQLEVGTVHLNNKTQRGTDNFPFLGAKKSGAGVQGVKYSIEAMTVKSVVFDIQ 475

A related GBS gene <SEQ ID 8815> and protein <SEQ ID 8816> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3  
 McG: Discrim Score: -15.11  
 GvH: Signal Score (-7.5): 0.17  
 Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 0 value: 1.22 threshold: 0.0  
 PERIPHERAL Likelihood = 1.22 187  
 modified ALOM score: -0.74

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2488(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

66.8/82.6% over 474aa

Streptococcus mutans  
 EGAD|42413| NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert characterized  
 EGAD|42413|110509 NADP-dependent glyceraldehyde-3-phosphate dehydrogenase Insert  
 characterized  
 SP|Q59931|GAPN\_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.9)  
 (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE  
 DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE). Edit characterized  
 GP|642667|gb|AAA91091.1|L38521 NADP-dependent glyceraldehyde-3-phosphate dehydro Insert  
 characterized

ORF01688(301 - 1725 of 2025)  
 EGAD|42413|44796(1 - 475 of 475) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase  
 {Streptococcus mutans|EGAD|42413|110509 NADP-dependent glyceraldehyde-3-phosphate  
 dehydrogenase {Streptococcus mutans|SP|Q59931|GAPN\_STRMU NADP-DEPENDENT GLYCERALDEHYDE-3-  
 PHOSPHATE DEHYDROGENASE (EC 1.2.1.9) (NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE  
 DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE [NADP+]) (TRIOSEPHOSPHATE  
 DEHYDROGENASE).GP|642667|gb|AAA91091.1|L38521 NADP-dependent glyceraldehyde-3-phosphate  
 dehydro  
 %Match = 49.3  
 %Identity = 66.7 %Similarity = 82.5  
 Matches = 317 Mismatches = 83 Conservative Sub.s = 75

195 225 255 285 315 345 375 405  
 \*GLKNLYFFTESLDIVKFLRKICQIIENR\*SDRINLLQCKRRFTLTKEYQNYVNGEWKSSVNQIETLSPIDSSSLGFVP  
 :||:|:||||| | | |:| | : || ||  
 MTKQYKNYVNGEWKLSENEIKIYEPASGAELGSVP  
 10 20 30

435 465 495 525 555 585 615 645  
 AMTREEVDHAMKAGREALPAWAALT VYERAQYLHKAADI IERDKKEIATVLAKKISKAYNASVTEVVRTADLIRYAAEEG  
 ||: |||: : :| ||| ||: ||| ||| ||: |||:| :|:|:| :| :|:| |||:| |||||  
 AMSTEEVDVYASAKKAQPAWRALS YIERAAYLHKVADILMRDKEKIGAILSKEVAKGYKSAVSEVVRTAEIINYAAEEG  
 50 60 70 80 90 100 110

[illegible]

GBS127-His was purified as shown in Figure 200, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1454

A DNA sequence (GBSx1540) was identified in *S.agalactiae* <SEQ ID 4465> which encodes the amino acid sequence <SEQ ID 4466>. Analysis of this protein sequence reveals the following:

```

Possible site: 17
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.37    Transmembrane    427 - 443 ( 427 - 443)

----- Final Results -----
        bacterial membrane --- Certainty=0.1150(Affirmative) < succ>
        bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:BAA78049 GB:AB027569 phosphoenolpyruvate-protein
phosphotransferase [Streptococcus bovis]
Identities = 534/577 (92%), Positives = 559/577 (96%)

Query: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAEEEARLDVALQASQDELSVIRE 60
MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDT+AEERLD AL+ASQDELS+IRE
Sbjct: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTSAEEEARLDAALKASQDELSIIRE 60

Query: 61 KAVESLGEEAAAVFDAHLMVLSDPEDINQIKETIRAKQVNAETGLKEVTDMFITIFEGME 120
KAVE+LGEERAAAVFDAHLMVL+DPEMI+QIKETIRAKQ NAE GLKEVTDMFITIFEGME
Sbjct: 61 KAVETLGEERAAAVFDAHLMVLADPEMISQIKETIRAKQTNAEAGLKEVTDMFITIFEGME 120
```

-1606-

5  
 Query: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180  
 DNPYMQERAADIRDVAKRVLAHLLG KLPNPATI+EESIVIAHDLTPSDTAQLNKQFVKA  
 Sbjct: 121 DNPYMQERAADIRDVAKRVLAHLLGAKLPNPATIDEESIVIAHDLTPSDTAQLNKQFVKA 180

10  
 Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDQQLIAVNGITGEVIIETPEAQISA 240  
 FVTNIGGRTSHSAIMARTLEIAAVLGTNDIT RV+DG ++AVNGITGEVII PT+ Q++  
 Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITSRVKDGDIVAVNGITGEVIINPTDEQVAE 240

15  
 Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAANIGTPKDVEGVNENGAEAVGLYRTE 300  
 FKAAGEAYAKQKAEWALLKDA+TVTADGKHFELAANIGTPKDVEGVN NGAEAVGLYRTE  
 Sbjct: 241 FKAAGEAYAKQKAEWALLKDAKTVTADGKHFELAANIGTPKDVEGVNANGAEAVGLYRTE 300

20  
 Query: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360  
 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPY DLPKEMNPFLGFR  
 Sbjct: 301 FLYMDSQDFPTEDEQYEAYKAVLEGMNGKPVVVRTMDIGGDKELPYLDLPKEMNPFLGFR 360

25  
 Query: 361 ALRISISSETGDMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKANLLAD 420  
 ALRISISSETG+AMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIF+EKANL A+  
 Sbjct: 361 ALRISISSETGNAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFDEEKANLKAE 420

30  
 Query: 421 GVAVAEGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSILYQP 480  
 GVAV++ I+VGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSILYQP  
 Sbjct: 421 GVAVSDDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSILYQP 480

35  
 Query: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQTA VPLLVMGLDEFMSATS+LRLRTRSL 540  
 YNPSILRLINNVIKAAHAEGKW GMCGEMAGDQ AVPLLVMGLDEFMSATS+LRLRTRSL  
 Sbjct: 481 YNPSILRLINNVIKAAHAEGKWVGMCGEMAGDQKAVPLLVMGLDEFMSATSILRTRSL 540

40  
 Query: 541 MKKLD TAKMEEYANRALSECSTMEEVIELQKEYVD F 577  
 MKKLD TAKM+EYANRAL+ECSTMEEV+EL KEYV+ D  
 Sbjct: 541 MKKLD TAKMQEYANRALTECSTMEEVLELSKEYVNVD 577

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4467> which encodes the amino acid  
 sequence <SEQ ID 4468>. Analysis of this protein sequence reveals the following:

Possible site: 39

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0875(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

45  
 Identities = 540/577 (93%), Positives = 561/577 (96%)

50  
 Query: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTNAEEARLDVALQASQDELSVIRE 60  
 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTV DTNAEEARLDVALQA+QDELSVIRE  
 Sbjct: 1 MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVADTNAEEARLDVALQAQDELSVIRE 60

55  
 Query: 61 KAVESLGEEAAAVFDAHLMVLSDPEDINQIKETIRAKQVNAETGLKEVTDMFITIFEGME 120  
 AVESLGEEAAAVFDAHLMVL+DPEMI+Q+KETIRAKQ NAETGLKEVTDMFITIFEGME  
 Sbjct: 61 NAVESLGEEAAAVFDAHLMVLADPEMISQVKETIRAKQTNAETGLKEVTDMFITIFEGME 120

60  
 Query: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180  
 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA  
 Sbjct: 121 DNPYMQERAADIRDVAKRVLAHLLGVKLPNPATINEESIVIAHDLTPSDTAQLNKQFVKA 180

65  
 Query: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITERVQDQQLIAVNGITGEVIIETPEAQISA 240  
 FVTNIGGRTSHSAIMARTLEIAAVLGTNDIT+RV+DG +IAVNGITGEVII+P+E Q+ A  
 Sbjct: 181 FVTNIGGRTSHSAIMARTLEIAAVLGTNDITRKVDGDIVAVNGITGEVIIDPSEDQVLA 240

70  
 Query: 241 FKAAGEAYAKQKAEWALLKDAQTVTADGKHFELAANIGTPKDVEGVNENGAEAVGLYRTE 300  
 FK AG AYAKQKAEW+LLKDA T TADGKHFELAANIGTPKDVEGVN+NGAEAVGLYRTE  
 Sbjct: 241 FKEAGAAYAKQKAESLLKDAHTETADGKHFELAANIGTPKDVEGVNDNGAEAVGLYRTE 300

-1607-

Query: 301 FLYMDSQDFPTEDEQEYAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360  
 FLYMDSQDFPTEDEQEYAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR  
 Sbjct: 301 FLYMDSQDFPTEDEQEYAYKAVLEGMNGKPVVVRTMDIGGDKELPYFDLPKEMNPFLGFR 360

5 Query: 361 ALRISISSETGDAMFRTQIRALLRASVHGQLRIMFPMVALLKEFRAAKAIFEEKANLLAD 420  
 ALRISISSETGDAMFRTQ+RALLRASVHGQLRIMFPMVALLKEFRAAKA+F+EEKANLLA+  
 Sbjct: 361 ALRISISSETGDAMFRTQMRALLRASVHGQLRIMFPMVALLKEFRAAKAVFDEEKANLLAE 420

10 Query: 421 GVAVAEGIEVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSYLYQP 480  
 GVAVA+ I+VGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSYLYQP  
 Sbjct: 421 GVAVADDIQVGIMIEIPAAAMLADQFAKEVDFFSIGTNDLIQYTMADRMNEQVSYLYQP 480

15 Query: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQTAVPLLVGMLDEFMSATSVLRTSL 540  
 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQ AVPLLVGMLDEFMSATSVLRTSL  
 Sbjct: 481 YNPSILRLINNVIKAAHAEGKWAGMCGEMAGDQQAVPLLVGMLDEFMSATSVLRTSL 540

20 Query: 541 MKKLD TAKMEEYANRAL SECSTMEEVIELQKEYVDFD 577  
 MKKLD+AKMEEYANRAL+ECST EEV+EL KEYV D  
 Sbjct: 541 MKKLSAKMEEYANRALTECSTAEEVLELSKEYVSED 577

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1455

25 A DNA sequence (GBSx1541) was identified in *S.agalactiae* <SEQ ID 4469> which encodes the amino acid sequence <SEQ ID 4470>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein is similar to a protein from *S.bovis*:

>GP:BAA78048 GB:AB027569 histidine containing protein [Streptococcus bovis]  
 Identities = 86/87 (98%), Positives = 87/87 (99%)

40 Query: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60  
 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD  
 Sbjct: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSIMGVMSLGVGQGAD 60

45 Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87  
 VTISAEGADADDA+AAIEETMTKEGLA  
 Sbjct: 61 VTISAEGADADDALAAIEETMTKEGLA 87

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4471> which encodes the amino acid sequence <SEQ ID 4472>. Analysis of this protein sequence reveals the following:

50 Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 86/87 (98%), Positives = 87/87 (99%)

-1608-

Query: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSLMGVMSLGVGGQAD 60  
 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSLMGVMSLGVGGQAD  
 Sbjct: 1 MASKDFHIVAETGIHARPATLLVQTASKFASDITLDYKGKAVNLKSLMGVMSLGVGGQAD 60

5 Query: 61 VTISAEGADADDAIAAIEETMTKEGLA 87  
 VTISAEGADA+DATAAIEETMTKEGLA  
 Sbjct: 61 VTISAEGADAEDAIAAIEETMTKEGLA 87

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 10 vaccines or diagnostics.

**Example 1456**

A DNA sequence (GBSx1542) was identified in *Sagalactiae* <SEQ ID 4473> which encodes the amino  
 acid sequence <SEQ ID 4474>. This protein is predicted to be glutaredoxin-like protein nrh (b2673).  
 Analysis of this protein sequence reveals the following:

15 Possible site: 56  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 20 bacterial cytoplasm --- Certainty=0.4532(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAA63372 GB:X92690 glutaredoxin-like protein [Lactococcus  
 lactis]  
 Identities = 42/70 (60%), Positives = 53/70 (75%)

Query: 4 ITVFSKNNCMQCKMTKKFLDQHGADFEEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63  
 +TV+SKNNCMQCKM KK+L +H F EINIDE+PE +E V +GF AAPVI + FS  
 30 Sbjct: 2 VTVYSKNNCMQCKMVKKWLSEHEIAFNEINIDEQPEFVEKVIEMGFRAAPVITKDDFAFS 61

Query: 64 GFQPSKLLKEL 73  
 GF+PS+L +L  
 35 Sbjct: 62 GFRPSELAKL 71

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4475> which encodes the amino acid  
 sequence <SEQ ID 4476>. Analysis of this protein sequence reveals the following:

Possible site: 17  
 40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4606(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 56/71 (78%), Positives = 68/71 (94%)

50 Query: 4 ITVFSKNNCMQCKMTKKFLDQHGADFEEINIDEKPEKIEYVKNLGFSAAPVIEAGNVVFS 63  
 ITV+SKNNCMQCKMTKKFL+QHG +F+EINIDE PEK++YVK+LGF++APVIEA N+VFS  
 Sbjct: 13 ITVYSKNNCMQCKMTKKFLDQHGAVNFQEIINIDEHPEKVDYVKSIGFTSAPVIEADNLVFS 72

Query: 64 GFQPSKLLKELV 74  
 GFQP+KLKEL+  
 55 Sbjct: 73 GFQPAKLLKELI 83

-1609-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1457

A DNA sequence (GBSx1543) was identified in *S.agalactiae* <SEQ ID 4477> which encodes the amino acid sequence <SEQ ID 4478>. This protein is predicted to be ribonucleotide reductase subunit R1E (nrDE). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3676(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD41036 GB:AF112535 ribonucleotide reductase alpha-chain

[Corynebacterium glutamicum]

Identities = 366/701 (52%), Positives = 488/701 (69%), Gaps = 19/701 (2%)

Query: 23 NGQIPLHKDKREALTAFFKENVQPNKAFDSITDKIAYLLKYDYLEEAFLNKYRPEFIEEL 82

NG+I KD+EA +F ++V N+ F ++ +KI YL++ Y + L+KY +FI++L

Sbjct: 22 NGKIQFEKDREAANQYFLOHVNQNTVFFHNLQEKIDYLVENKYYDPIVLDKYDFQFIKDL 81

Query: 83 STKLFDDKKFRKSFMAAYKFYQYALKTNDGEYILESIDRVLFNALYFADGDEELATDL 142

+ + KFRF+SF+ AYK+Y Y LKT DG YLE EDRV AL ADGD LA +L

Sbjct: 82 FKRAYGFKFRFQSFLGAYKYYTSYTLKTFDGRRYLERFEDRVCMVALTLADGDRALENL 141

Query: 143 ALEMISQRYQPATPSFLNAGRSRRGELVSCFLIQVTDDMNAIGRSINSALQLSRIGGGVG 202

E++S R+QPATP+FLN+G+++RGE VSCFL+++ D+M +IGRSINSALQLS+ GGGV

Sbjct: 142 VDEIMSGRFQPATPTFLNSGKAQRGEFVSCFLLRIEDNMESIGRSINSALQLSKRGGGVA 201

Query: 203 ISLSNLREAGAPIKGFAGAASGVVPMKLFEDSFSYSNQLGQRQAGVVYLDVPHDIIIS 262

+ LSNLREAGAPIK +SGV+PVMKL ED+FSY+NQLG RQGAG VYL+ HPDI+S

Sbjct: 202 LLLSNLREAGAPIKKIENQSSGVIPVMKLLLEDASFANQLGARQAGAVYLNAAHHPDILS 261

Query: 263 FLSTKKENADEKVRVKTLISLGITVPDKFYELARNNQEMYLFSPYSIEREYGVFPFSYIDIT 322

FL TK+ENADEK+R+KTLISLG+ +PD +ELA+ N +MYLFSPY +ER YG PF+ + IT

Sbjct: 262 FLDTKRENADEKIRIKTISLGVVPIPDITFELAKRNDMYLFSPYDVERIYKPFADVST 321

Query: 323 EKYDELVANPNITKTKINARDLETEISKLQQESGYPIIINIDTANRTNPVDGKIIMSNLC 382

E YDE+V + I KTKINAR +++++Q ESGYPYI+ DT N +NP++G+I SNLC

Sbjct: 322 EHYDEMVDDDRIRKTKINARQFFQTLAEIQFESGYPYIMYEDTVNASNPIEGRITHSNLC 381

Query: 383 SEILQVQKPSLINDAQEYLEMGTDISCNLGSSTNVNMMTSPDFGKSIKTMTRALTFTVDS 442

SEILQV PS ND Y E+G DISCNLGS NV M SP+F K+I+T R LT V++

Sbjct: 382 SEILQVSTPSEFNDDLTYAEVGEDISCNLGSINLVAMAMDSPNFECTIETAIRGLTAVSEQ 441

Query: 443 SNIEAVPTIKNGNAQAHTFGLGAMGLHSYLAKNHIEYGSPESIEFTDIYFMLMNYWTLVE 502

++I++VP+I+ GN AH GLG M LH Y + H+ YGS E+++FT+ YF + Y L

Sbjct: 442 TSIDSVPISIRKGNAAHAIGLQGMNLHGYFGREHMHYGSEALDFTNAYFAAVLYQCLRA 501

Query: 503 SNNIARERQTTTFVGFEKSKYADGTYFDKYVSGKFVPQSDKVKSLFA--NHFIPEAKDWEN 560

SN IA ER F FE SKYA G YFD + + F P+SDKVK LFA N P +DW

Sbjct: 502 SNKIATERGERFKNFENSKYATGEYFDDFDANDFAPKSDKVKELFAKSNIHTPTVEDWAA 561

Query: 561 LRYAVMKDGLYHQNRLAVAPNGSISYINDCSASIHPIQRIEERQEKKIGKIYYPANGLA 620

L+ VM+ GL+++N AV P GSISYIN+ ++SIHPI +IE R+E KIG++YYP A +

Sbjct: 562 LKADVMEHGLFNRNLQAVPPTGSISYINNSTSSIHPIASKIEIRKEGKIGRVYPAPHMD 621

Query: 621 TDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLRSELPKELYEWKTESKQTTRD 680

D + Y+ AY++ K+ID YA AT++VDQGLS+TLF + TTRD

Sbjct: 622 NDNLEYFEDAYEIGYEKIIDTYAVATKYVDQGLSLTLFFK-----DTATTRD 668

-1610-

Query: 681 LSILRNAYAFNKGVKSIYYI--RTFTDDGSEVGANQCESCVI 719  
 ++ + YA+ KG+K++YYI R +G+EV + C SC++  
 Sbjct: 669 INRAQIYAWRKGIKTLYYIRLRQVALEGTEV--DGCVSCML 707

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4479> which encodes the amino acid sequence <SEQ ID 4480>. Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 628/719 (87%), Positives = 682/719 (94%)

Query: 1 MSLKNIGDVSYFRLNNEINRPVNGQIPLHKDKREALTAFFKENVQPSKAFDSITDKIAYL 60  
 MSLK++GD+SYFRLNNEINRPVNG+IPLHKDKREAL AF ENV PN+ +F SIT+KI YL  
 Sbjct: 1 MSLKDLGDISYFRLNNEINRPVNGKIPLHKDKREALKAFSAENVLPNTMSFTSITEKIEYL 60

Query: 61 LKYDYLEEAFLNKYRPEFIEELSTKLFDKKFRFKSFMAAYKFYQQYALKTNDGEYYLESI 120  
 + DY+E AF+ KYRPEFI EL + + + FRFKSFMAAYKFYQQYALKTNDGE+YLE++  
 Sbjct: 61 ISNDYIESAFIQKYRPEFITELDSIIKSENFRFKSFMAAYKFYQQYALKTNDGEHYLENL 120

Query: 121 EDRVLFNALYFADGDEELATDLALEMISQRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180  
 EDRVLFNALYFADG E+LA DLA+EMI+QRYQPATPSFLNAGRSRRGELVSCFLIQVTDD  
 Sbjct: 121 EDRVLFNALYFADGQEDLAKDLAVEMINQRYQPATPSFLNAGRSRRGELVSCFLIQVTDD 180

Query: 181 MNAIGRSINSALQLSRIGGGVGISLSNLREAGAPIKGFGAASGVVPVMKLFEDSFSYSN 240  
 MN+IGRSINSALQLSRIGGGVGI+LSNLREAGAPIKG+AGAASGVVPVMKLFEDSFSYSN  
 Sbjct: 181 MNSIGRSINSALQLSRIGGGVGITLSNLREAGAPIKGYAGAASGVVPVMKLFEDSFSYSN 240

Query: 241 QLGQRQAGVVYLDVPHPDIIISFLSTKKENADEKVRVKTLISLGITVPDKFYELARNNQEM 300  
 QLGQRQAGVVYL+VFHPDII+FLSTKKENADEKVRVKTLISLGITVPDKFYELAR N++M  
 Sbjct: 241 QLGQRQAGVVYLVNPFHPDIIAFLSTKKENADEKVRVKTLISLGITVPDKFYELARKNEDM 300

Query: 301 YLFSPYSIEREYGVPPFSYIDITEKYDELVANPNITTKINARDLETEISKLQQESGYPI 360  
 YLFSPY++E+EYG+PF+Y+DIT YDELVANP ITTKI ARDLETEISKLQQESGYPI  
 Sbjct: 301 YLFSPYNVEKEYGIPFNYLDITNMYDELVANPKITTKIKARDLETEISKLQQESGYPI 360

Query: 361 INIDTANRTNPVDGKIIMSNLCSEILQVQKPSLINDAQEYLEMGTDISCNLGSTNVLNMM 420  
 INIDTAN+ NP+DGKIIMSNLCSEILQVQ PSLINDAQE++EMGTDISCNLGSTN+LNMM  
 Sbjct: 361 INIDTANKANPIDGKIIMSNLCSEILQVQTPSLINDAQEFVEMGTDISCNLGSTNINMM 420

Query: 421 TSPDFGKSIKTMTRALTFVTDSSNIEAVPTTIKNGNAQHTFGLGAMGLHSYLAKNHIEYG 480  
 TSPDFG+SIKTMTRALTFVTDSS+IEAVPTTIK+GN+QAHTFGLGAMGLHSYLA++HIEYG  
 Sbjct: 421 TSPDFGRSIKTMTRALTFVTDSSSIEAVPTTIKHGNSQAHTFGLGAMGLHSYLAQHIEYG 480

Query: 481 SPESIEFTDIYFMLMNYWTLVESNNIARERQTTFVGFEKSYADGTYFDKYVSGKFVPQS 540  
 SPESIEFTDIYFML+NYWTLVESNNIARERQTTFVGFE SKYA+G+YFDKYV+G FVP+S  
 Sbjct: 481 SPESIEFTDIYFMLLNWTLVESNNIARERQTTFVGFEKSYANGSYFDKYVTGHFVPKS 540

Query: 541 DKVKSLFANHFIPEAKDWENLRYAVMKDGLYHQNRLAVAPNGSISYINDCSASIHPIQR 600  
 D VK LF +HFIP+A DWE LR AV KDGLYHQNRLAVAPNGSISYINDCSASIHPIQR  
 Sbjct: 541 DLVKDLFKDHFIPQASDWEALRDAVQKDGLYHQNRLAVAPNGSISYINDCSASIHPIQR 600

Query: 601 IEERQEKKIGKIYYPANGLATDTIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSMTLFLR 660  
 IEERQEKKIGKIYYPANG+TDITIPYYTSAYDMDMRKVIDVYAAATEHVDQGLS+TLFLR  
 Sbjct: 601 IEERQEKKIGKIYYPANGSLDTITIPYYTSAYDMDMRKVIDVYAAATEHVDQGLSLTLFLR 660

Query: 661 SELPKELYEWKTESKQTTRDLSILRNAYAFNKGVKSIYYIRTFTDDGSEVGANQCESCVI 719  
 SELP ELYEWT+SKQTTRDLSILRNAYAFNKG+KSIYYIRTFTDDG EVGANQCESCVI  
 Sbjct: 661 SELPMELYEWKTQSKQTTRDLSILRNAYAFNKGKSIYYIRTFTDDGEEVGANQCESCVI 719

-1611-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1458

A DNA sequence (GBSx1544) was identified in *S.agalactiae* <SEQ ID 4481> which encodes the amino acid sequence <SEQ ID 4482>. This protein is predicted to be ribonucleotide reductase subunit R2F (nrdB). Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9753> which encodes amino acid sequence <SEQ ID 9754> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAC14561 GB:AF050168 ribonucleoside diphosphate reductase small
subunit [Corynebacterium ammoniagenes]
Identities = 166/313 (53%), Positives = 215/313 (68%), Gaps = 1/313 (0%)

Query: 10 EAINWNEIEDVIDKSTWEKLTQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFGGLTLL 69
      +AINWN I D D W++LT FWL +IP+SND+ W K++ QE+ +VF GLTLL
Sbjct: 17 KAINWNVIPDEKDLLEVWDRLTGNFWLPEKIPVSNDIQSWNKMTPOEQLATMRVFTGLTLL 76

Query: 70 DTMQSETGVEAIRADVTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEIFEWTN 129
      DT+Q G ++ DV T HEE V NI FMESVHAKSYS+IF TL + +I E F W+
Sbjct: 77 DTIQGTVGAISSLPLDVTMHEEGVYTNI AFMESVHAKSYSNIFMTLASTPQINEAFRWSE 136

Query: 130 NNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYYLGNKLANVAEIIK 189
      NE LQ KA+II Y + L+KKVAST LE+FLFYSGF+ P+Y KL N A+II+
Sbjct: 137 ENENLQRKAKIIMSYYNGDDPLKKKVASTLLESFLFYSGFYLP MYLSSRAKLINTADIIR 196

Query: 190 LIIRDES VHGTIYGYKQFGFNLPEDEQENFRDWMYDLLYQLYENEEKYTKTLYDGVGW 249
      LIIRDES VHG YIGYK+Q G +L E EQE ++ + +DL+Y LYENE +YT+ +YD +GW
Sbjct: 197 LIIRDES VHGYIYGYKQGGVKKLSEAEQEYKAYTFDLMYDLYENEIEYTEDIYDDL GW 256

Query: 250 TEEVMTFLRYNANKALMNLGQDPLFPDPTANDVNPIVMNGIS-TGTSNHDFFSQVNGGYLL 308
      TE+V FLRYNANKAL NLG + LFP V+P +++ +S NHDFFS G+ Y++
Sbjct: 257 TEDVKRFLRYNANKALNNLGYEGLFPTDETKVSPAILSSLSPNADENHDFFSGSGSSYVI 316

Query: 309 GSVEAMHDDDDYNY 321
      G E DDD+++
Sbjct: 317 GKAE DTTDDDDWDF 329
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4483> which encodes the amino acid sequence <SEQ ID 4484>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4583(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

-1612-

Identities = 315/319 (98%), Positives = 316/319 (98%)

Query: 5 MTTYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFG 64  
 5 Sbjct: 1 MTTYEAINWNEIEDVIDKSTWEKLTEQFWLDTRIPLSNDLDDWRKLSAQEKDLVGKVFG 60

Query: 65 GLTLDDTMQSETGVEAIRADVTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEI 124  
 10 Sbjct: 61 GLTLDDTMQSETGVEAIRADVTPHEEAVLNNIQFMESVHAKSYSSIFSTLNTKSEIEEI 120

Query: 125 FEWTNNNEFLQEKARIINDIYANGNALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANV 184  
 Sbjct: 121 FEWTNNNEFLQEKARIINDIYANGDALQKKVASTYLETFLFYSGFFTPLYYLGNNKLANV 180

Query: 185 AEIIKLIIRDES VHGTIYGK FQLG FNELPEDEQENFRDWMYD LLYQLYENEEKYTKTLY 244  
 15 Sbjct: 181 AEIIKLIIRDES VHGTIYGK FQLG FNELPEDEQENFRDWMYD LLYQLYENEEKYTKTLY 240

Query: 245 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIMNGISTGTSNHDFFSQVGN 304  
 20 Sbjct: 241 DGVGWTEEVMTFLRYNANKALMNLGQDPLFPDTANDVNPIMNGISTGTSNHDFFSQVGN 300

Query: 305 GYLLGSVEAMHDDDDYNYGL 323  
 25 Sbjct: 301 GYLLGSVEAMHDDDDYNYGL 319

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1459**

30 A DNA sequence (GBSx1545) was identified in *S.agalactiae* <SEQ ID 4485> which encodes the amino acid sequence <SEQ ID 4486>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence  
 35 INTEGRAL Likelihood = -0.27 Transmembrane 50 - 66 ( 50 - 66)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 40 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1460**

A DNA sequence (GBSx1546) was identified in *S.agalactiae* <SEQ ID 4487> which encodes the amino acid sequence <SEQ ID 4488>. Analysis of this protein sequence reveals the following:

Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 50 INTEGRAL Likelihood = -14.38 Transmembrane 176 - 192 ( 168 - 201)  
 INTEGRAL Likelihood = -4.57 Transmembrane 25 - 41 ( 22 - 42)  
 INTEGRAL Likelihood = -3.88 Transmembrane 94 - 110 ( 94 - 112)  
 INTEGRAL Likelihood = -1.49 Transmembrane 70 - 86 ( 70 - 86)  
 55 INTEGRAL Likelihood = -1.01 Transmembrane 128 - 144 ( 128 - 144)

----- Final Results -----

-1613-

bacterial membrane --- Certainty=0.6753 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 5 A related GBS nucleic acid sequence <SEQ ID 9751> which encodes amino acid sequence <SEQ ID 9752> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 55/184 (29%), Positives = 98/184 (52%), Gaps = 4/184 (2%)

Query: 16 MSKNNNTTCLTETAIFAALAMALSMIP---DFASWFTPSFGAIPILFALRRGTKYGLF 71  
 M+++ LIE AI A A+ L ++ + S IP+ L + R G K GL  
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIVSGMFLSMPQGGSVSIMMIPIFLISFRWGVKAGLT 60

Query: 72 AGLIWGLLHFVLSKVYYLSLSQVFIIEYILAFISMGLAGVFSKFKDALSSSSSKTKALSIA 131  
 GL+ GL+ + ++ Q+ ++YI+AF ++G++G F++ + A S +K K +  
 Sbjct: 61 TGLLTGLVQIAIGNLFAQHPVQLLLDYIVAFAAIGISGCFASSVRKAASVSKTKGKLIVSV 120

Query: 132 LSGAILATLVRYVWHYIAGVIFWASYAPKGMSATLYSLSVNGTAGLLTLFFVVISIILV 191  
 +S + +L+RY H I+G +F+ S+APKG +YSL+ N T + + I + +L  
 Sbjct: 121 VSAVFIGSLLRYAAHVISGAVFFGSPAPKGTVPVWIYSLTYNATYMPVSFIICAIVLCLLF 180

Query: 192 ISYP 195  
 ++ P

Sbjct: 181 MTAP 184

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4489> which encodes the amino acid sequence <SEQ ID 4490>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -9.34 Transmembrane 162 - 178 ( 156 - 183)  
 INTEGRAL Likelihood = -9.34 Transmembrane 110 - 126 ( 107 - 130)  
 INTEGRAL Likelihood = -1.22 Transmembrane 55 - 71 ( 55 - 71)

----- Final Results -----

bacterial membrane --- Certainty=0.4736 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 40 The protein has homology with the following sequences in the databases:

>GP:CAB15077 GB:Z99119 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 55/189 (29%), Positives = 100/189 (52%), Gaps = 10/189 (5%)

Query: 1 MSPNTNVKYLIEAAIFAALAMTSLFIPDFAGWF--SPSYGAIALV-----IFSLRRGLKY 53  
 M+ + + LIE AI A A+ L + +G F P G++++ + S R G+K  
 Sbjct: 1 MNQSKQLVRLIEIAIMTAAAVILDIV---SGMFLSMPQGGSVSIMMIPIFLISFRWGVKA 57

Query: 54 GMLAGLIWGLLHFVLGKVYYLSMSQVFIIEYILAFISMGLAGSFSDSLIKTLRRQQTFFAV 113  
 G+ GL+ GL+ +G ++ Q+ ++YI+AF ++G++G F+ S+ K + +  
 Sbjct: 58 GLTTGLLTGLVQIAIGNLFAQHPVQLLLDYIVAFAAIGISGCFASSVRKAASVSKTKGKLI 117

Query: 114 FLAIMASLLAVTVRYLWHFLAGIIFWGSYAPKGMSAVWYSFVNGTAGVLTFLITCLALM 173  
 + A + +RY H ++G +F+GS+APKG YS + N T V +F+I + L  
 Sbjct: 118 VSVVSAVFIGSLLRYAAHVISGAVFFGSPAPKGTVPVWIYSLTYNATYMPVSFIICAIVLC 177

Query: 174 IALPIHPQL 182  
 + P+L

Sbjct: 178 LLFMTAPRL 186

- 60 An alignment of the GAS and GBS proteins is shown below.

Identities = 116/186 (62%), Positives = 138/186 (73%)

-1614-

Query: 16 MSKNNNTTCLIFETAIFAALAMALSMIPDFASWFTPSFGAIPILFALRRGTYGLFAGLI 75  
 MS N N LIE AIFAALAM LS IPDFA WF+PS+GAI L++F+LRRG KYG+ AGLI  
 Sbjct: 1 MSPNTNVKYLIEAAIFAALAMTLSFIPDFAGWFSYGAIALVIFSLRRLKYGMLAGLI 60

Query: 76 WGLLHFVLSKVYYLSLSQVFIEYILAFISMGLAGVFSAKFKDALSSSSKTKALSALSGA 135  
 WGLLHFVL KVYYLS+SQVFIEYILAF SMGLAG FS L A+ LA+ +  
 Sbjct: 61 WGLLHFVLGKVYYLSMSQVFIEYILAFISMGLAGSFSDSLIKTLRRQQTFFAVFLAIMAS 120

Query: 136 ILATLVRYVWHYIAGVIFWASYAPKGSATLYSLSVNGTAGLLTLFFVVISIILVISYP 195  
 +LA VRY+WH++AG+IFW SYAPKMSA YS SVNGTAG+LT ++++I + +P  
 Sbjct: 121 LLAVTVRYLWHFLAGIIFWGSYAPKGSASVWYSFVNGTAGVLTFLITCLALMIALPIHP 180

Query: 196 SFFLPK 201  
 F PK  
 Sbjct: 181 QLFDPK 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1461

A DNA sequence (GBSx1547) was identified in *S.agalactiae* <SEQ ID 4491> which encodes the amino acid sequence <SEQ ID 4492>. Analysis of this protein sequence reveals the following:

Possible site: 34

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -7.43	Transmembrane	206 - 222 ( 199 - 223)
INTEGRAL	Likelihood = -6.64	Transmembrane	24 - 40 ( 19 - 42)
INTEGRAL	Likelihood = -6.58	Transmembrane	61 - 77 ( 51 - 78)
INTEGRAL	Likelihood = -6.58	Transmembrane	134 - 150 ( 132 - 154)
INTEGRAL	Likelihood = -4.62	Transmembrane	226 - 242 ( 224 - 245)
INTEGRAL	Likelihood = -3.72	Transmembrane	107 - 123 ( 106 - 125)

----- Final Results -----

bacterial membrane --- Certainty=0.3972(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9749> which encodes amino acid sequence <SEQ ID 9750> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4493> which encodes the amino acid sequence <SEQ ID 4494>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.46	Transmembrane	134 - 150 ( 131 - 159)
INTEGRAL	Likelihood = -7.59	Transmembrane	107 - 123 ( 103 - 128)
INTEGRAL	Likelihood = -7.48	Transmembrane	225 - 241 ( 213 - 248)
INTEGRAL	Likelihood = -7.22	Transmembrane	205 - 221 ( 199 - 224)
INTEGRAL	Likelihood = -3.56	Transmembrane	50 - 66 ( 50 - 73)
INTEGRAL	Likelihood = -1.28	Transmembrane	16 - 32 ( 16 - 33)

----- Final Results -----

bacterial membrane --- Certainty=0.5182(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1615-

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/253 (32%), Positives = 149/253 (58%), Gaps = 5/253 (1%)

```

5  Query: 6  IKQSDTTFVRIIKSLIGGFIGAILGSGALFIIF--GQDKYLSEI--NIVQYFLWVSRI 61
      +K+  +F+R++K  L+   G I+G +   F+ +   G+  +L+ +   +++  + ++R+
Sbjct: 1  MKKKKNSFLRLKMSLLSSLAGGIIGMVGAFILGYHGGRLDHLTFLKDDVINLIILLNRL 60

10 Query: 62  VVIITALFSLIYLYQIQKYQKVFFNVDESQ--SEEIYRQINLRHSYGMTFVSISIVLSIVN 120
      VV+   S ++L Q++K  V+  ++E   SE YRQ+N +H+Y M  ++++ +LS+ N
Sbjct: 61  VVVTDLTLSFVFLTQLKKEAVYNTIEEDDISSENGYRQLNKKHAYTMLLIASILSMCN 120

15 Query: 121 TLFNYKLNIFDSDSVTLVIPIYDLSLLFVLLGLHIYFLKVYRNIRGIKMTVAPTLLKELKNN 180
      L   L           L IP+ D+ LL +++          +K Y  IRG  +   P LKELK+N
Sbjct: 121 VLLGLTLTNDSQHAMLAIPLLDILLLLMVIPFQALAMKRYNAIRGTDVPYFPNLKELKHN 180

20 Query: 181 VLQLDEAELESNYKMCFDIVMNLSGFIFPTIYFVLFFISFVFQKVEIVAIITTSIHIYI 240
      ++ LDEAEL++ +K  F+ V++L+G I P++Y +LFF+          +VE+ AI++   I +Y+
Sbjct: 181 IMALDEAELQAYHKTSFESVLSLNGVVIIPSLYVILFFVYLFTGQVELTAILVLVLIQLYL 240

25 Query: 241 LIKSLKAARHFYR 253
      L+KS      R FYR
Sbjct: 241 LVKSATMTRQFYR 253

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1462

A DNA sequence (GBSx1548) was identified in *S.agalactiae* <SEQ ID 4495> which encodes the amino acid sequence <SEQ ID 4496>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 57
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.5172(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1463

A DNA sequence (GBSx1549) was identified in *S.agalactiae* <SEQ ID 4497> which encodes the amino acid sequence <SEQ ID 4498>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 26
    >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2059(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC76650 GB:AE000440
      UDP-D-glucose: (galactosyl) lipopolysaccharide

```

-1616-

glucosyltransferase [Escherichia coli K12]

Identities = 70/256 (27%), Positives = 121/256 (46%), Gaps = 14/256 (5%)

5 Query: 1 MNLFSIDDMYVDHFKVMLYSLVRQTKNRKLEIYVLQKT----LLKRHTELIQYTONLEV 56  
 +N+ + +D Y+D V + S+V ++ L+ Y++ ++ +L + Q  
 Sbjct: 28 LNVAYGVDANYLDGVGVSITSIVLNNRHINLDFYIITADVYNDGFFQKIAKLAEQNQLRIT 87

10 Query: 57 GYHPPIIVGTEVFAQAPTITDRYPDTIYRLLAHKFLPETLDRILYLDADMLCLNDFSSLYD 116  
 Y + T+ P T + +Y+RL A + L TLDR+LYLDAD++C D S L  
 Sbjct: 88 LYR---INTDKLQCLPCTQVWSRAMYFRLFAFQLLGLTLDRLLYLDADVVCKGDISQLLH 144

15 Query: 117 MELGDQLYAAASHNTDGKFLDYVNKLRLKNVELESSYFNTGVLLMNLPAIRKVVHQQTIL 176  
 + L A A+ D + + RL + EL YFN+GV+ ++L + L  
 Sbjct: 145 LGLNG---AVAAVVKDVEPMQEKAVSRSLSDPELLGQYFNSGVVYLDLKKWADAKLTEKAL 201

20 Query: 177 DYIMQNRGRLILPDQDILNGLYANLVKPIPDDEITYNYDARYSLIYQLKSRNEWDLEWVINH 236  
 +M PDQD++N L + +P E Y+ Y++ +LK + + +I  
 Sbjct: 202 SILMSKDNVYKYPDQDVMNVLLKGMTLFLPRE---YNTIYTIKSELKDKTHQNYKKLITE 258

20 Query: 237 -TVFLHFAGRDKPWKK 251  
 T+ +H+ G KPW K  
 Sbjct: 259 STLLIHYTGATKPWHK 274

No corresponding DNA sequence was identified in *S.pyogenes*.

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1464**

A DNA sequence (GBSx1550) was identified in *S.agalactiae* <SEQ ID 4499> which encodes the amino acid sequence <SEQ ID 4500>. Analysis of this protein sequence reveals the following:

30 Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 35 bacterial cytoplasm --- Certainty=0.1406(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1465**

A DNA sequence (GBSx1551) was identified in *S.agalactiae* <SEQ ID 4501> which encodes the amino acid sequence <SEQ ID 4502>. Analysis of this protein sequence reveals the following:

45 Possible site: 54  
 >>> Seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -10.72 Transmembrane 7 - 23 ( 1 - 28)  
 INTEGRAL Likelihood = -4.30 Transmembrane 222 - 238 ( 216 - 238)  
 INTEGRAL Likelihood = -3.66 Transmembrane 151 - 167 ( 140 - 170)  
 INTEGRAL Likelihood = -3.50 Transmembrane 35 - 51 ( 34 - 58)  
 INTEGRAL Likelihood = -3.35 Transmembrane 71 - 87 ( 69 - 88)  
 INTEGRAL Likelihood = -3.29 Transmembrane 113 - 129 ( 113 - 132)  
 INTEGRAL Likelihood = -2.81 Transmembrane 170 - 186 ( 168 - 190)  
 INTEGRAL Likelihood = -2.71 Transmembrane 198 - 214 ( 197 - 217)

-1617-

## ----- Final Results -----

bacterial membrane --- Certainty=0.5288(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07774 GB:AP001520 unknown conserved protein [Bacillus halodurans]  
 Identities = 84/242 (34%), Positives = 147/242 (60%), Gaps = 16/242 (6%)

Query: 1 MVGLGTVINVILIIVGGFVGLFLKNFLKESLQKSLMQAMGVAVLFISISGVLEKMLLVEK 60  
 MV +GTV+N I++ +GL +KN + E ++ +LMQA+G+A++ + + KM L +  
 Sbjct: 1 MVLIGTVVNGAAIVIAALIGLLVKN-IPERVKTTLMQAIGLAIVLLGV-----KMGLQTE 54

Query: 61 SHLISNHTNMIIITLALGTVLGELLSDSYIDKFGNYLKQKTGSGNDIKFVEAFVTSTCT 120  
 LI +I +L +G V+GE+++L+ +D G +++ K G D AFVT+T  
 Sbjct: 55 QFLI-----VICSLVIGGVIGEMINLEKRLDHLGRWIESKVGGKKDGSIAATAFVTTTLI 108

Query: 121 VCIGAMAVVGSIQDGIADHSILFAKGLDMIIIAIMTVSLGKALFSALPVALQGS LT 180  
 +GAMAV+G++ G+ DHS+L K +LD + + T +LG G LFSA+PV L QGS+  
 Sbjct: 109 YVVGAMAVLGAIDSGLRGDHSLVLLTKALLDGLAILFTSTLGIGVLFSAIPVVLYQGSIA 168

Query: 181 IVAF----FMGSLNPNSSLDYLNVLGNMLIFCVGVNLLFNLNKVINMLPAIILAILWGS 236  
 + A ++ + L S + ++ G ++I +G+NLL +NI+V N+LP++++ + +  
 Sbjct: 169 LFASQIDQYVPTALMDSFITMSATGGVMIVAIGLNLNVNIRVANLLPSLVIVAVLVT 228

Query: 237 FI 238  
 F+

Sbjct: 229 FV 230

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1466**

A DNA sequence (GBSx1552) was identified in *S.agalactiae* <SEQ ID 4503> which encodes the amino acid sequence <SEQ ID 4504>. This protein is predicted to be alanyl-tRNA synthetase (alaS). Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.41 Transmembrane 805 - 821 ( 804 - 822)

## ----- Final Results -----

bacterial membrane --- Certainty=0.2763(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04986 GB:AP001511 alanyl-tRNA synthetase [Bacillus halodurans]  
 Identities = 482/885 (54%), Positives = 618/885 (69%), Gaps = 27/885 (3%)

Query: 1 MKELSSAQIRQMWLDFWKSKGHSVEPSANLVPVNDPTLLWINSVATLKKYFDG SVIPEN 60  
 MK L+SAQ+RQM+LDF+K KGH VEPSA+LVP +DP+LLWINSVATLKKYFDG VIPEN  
 Sbjct: 1 MKYLTSAQVRQMFLDFFKEKGHDVEPSASLVPHDDPSLLWINSVATLKKYFDGRVIPEN 60

Query: 61 PRITNAQKSIRTNDIENVGKTARHHTFMELGNFSIGDYFRDEAIEWGFELLTSPEW FDF 120  
 PRITNAQKSIRTNDIENVGKTARHHT FEMLGNFSIGDYF++EAIEW +E LTS +W F  
 Sbjct: 61 PRITNAQKSIRTNDIENVGKTARHHTFFEMLGNF SIGDYFKEEAIEWAWEFLTSEKWIGF 120

Query: 121 PKDKLYMTYYPDDKDSYNRWIA-CGVEPSHLVPIEDNFWIEGAGPSGPDTEIFFDRGEDF 179

-1618-

K+KL +T +P+D ++Y+ W G+ ++ +E NFW+IG GPSGP+TEIF+DRG ++  
 Sbjct: 121 DKEKLSVTVHPEDDEAYSYWKEKIGIPEERIIRLEGNFWDIGEGPSGPNTTEIFYDRGPPEY 180

5 Query: 180 -----DPENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGL 234  
 DPE L ENDRY+E+WN+V SQFN +P Y LP KNIDTG GL  
 Sbjct: 181 GDQPNDE-----LYPGGENDRYLEVWNLVFSQFNHNPD---GSYTPLEPKKNIDTGMGL 231

10 Query: 235 ERLAAVMQGAKTNFETDLFMPPIIREVEKLSGKTYDPDGD-NMSFKVIADHIRALSFAIGD 293  
 ER+ +V+Q TNFETDLFMPPIIR EK+SG Y + ++SFKVIADHIR ++FAIGD  
 Sbjct: 232 ERMVSVIQNVPTNFETDLFMPPIIRATEKISGTEYGSHEADVSPFKVIADHIRTVTFAIGD 291

15 Query: 294 GALPGNEGRGYVLRRLRRRAVMHGRRLGINETFLYKLVPTVVGQIMESYYPEVLEKRDIFIE 353  
 GALP NEGRGYVLRRLRRRAV + +++GI+ F+Y+LVP VG IM +YPEV EK FI+  
 Sbjct: 292 GALPSNEGRGYVLRRLRRRAVRYAKQIGIDRPFMYELVPVVGIMVDFYPEVKEKAFFIQ 351

20 Query: 354 KIVKREETPARTIDAGSGHLDLSLAQLKAEGKDTLEGKDIFKLYDTYGFVPELTELAE 413  
 K+VK EEE F T++ G L+ ++ + K+EG T+ G D+F+LYDTYGFV+LTEE E  
 Sbjct: 352 KVVKTEERFHETLNEGLSILEKVIDKAKSEGASTISGSDVFRLYDTYGFVVDLTEEYVE 411

25 Query: 414 DAGYKIDHEGFSKAMKEQQDRARAADVKGSGMGMQNETLAGIVEESRF-EYDYSLESSL 472  
 + G ++D +GF++ M+ Q++RAR A + GSM +Q+E L I +S F Y S E+++  
 Sbjct: 412 BQGLQVLDLGFEMERQREARARARQAGSMQVQDEVLGQITVDSTFIGYKQLSTETTI 471

30 Query: 473 SVIADNERTEAVSEGO-ALLVFAQTFFYAEMGGQVADHGVKNDKGDVAVVVDVQKAP 531  
 I+ D + V GQ A ++ +TPFYAE GGQVAD G+I+ G V V DVQKAP  
 Sbjct: 472 ETIVLDKTVADYVAGAGQEAQVILKETPFYAESGGQVADKGIIRGANGFAV--VSDVQKAP 529

35 Query: 532 NGQPLHTVNVL-ASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHNHVICEHATQAGS 590  
 NGQ LHTV V +L V + + R + KNHTATHLLH AL +V+GEH QAGS  
 Sbjct: 530 NGQHLHTVIVKEGTLQVNDQVQAIVEETERSGIVKNHTATHLLHRALKDVLGEHVNQAGS 589

40 Query: 591 LNEEEFLRFDFTHFEAVSNEELRHIEQEVNEQIWNDLTITTTTETDVETAKEMGAMALFGE 650  
 L EE LRFDH+HF V++EE IE+ VNE+IW + + + ++ AK +GAMALFGE  
 Sbjct: 590 LVSEERLRFDFSHFGQVTDEEKEKIERIVNEKIWAQKVNISTKTLDEAKAIGAMALFGE 649

45 Query: 651 KYGKVVVVQIGNYSVELCGGTHLNNSSSEIGLFKIVKEEGIGSGTRRIIAVTGRQAFAY 710  
 KYG +VRVV++G+YS+ELCGG H+ N+SEIGLFKIV E GIG+G RRI AVTG++AF  
 Sbjct: 650 KYGDIVRVVEVGDISIELCGGCHVTNTSEIGLFKIVSESGIGAGVRRIEAVTGKEAFLFM 709

50 Query: 711 RNQEDALKEIAATVKAPQLKDAAAKQVQALSDSLRDLQKENVELKEAAAAAGDVFKDQI 770  
 Q D LKE AATVKA +KD +V+AL +R+LQ+EN L K AG + ++Q  
 Sbjct: 710 AKQLDLLKETAATVKAKNVKDVPRVEALQQQIRELQRENESLNKLGNEAGSLVNEVQ 769

Query: 771 EAKGVRFIASQVDVADAGALRTFADNWKQKDYSDVLVLAAGEKVNVLVASKTKDV--- 827  
 + +GV +A + AD LR+ D KQ+ S V+VL A KVN+ VA TKD+  
 Sbjct: 770 KIEGVPVLAKAISGADMDGLRSIVDKLKQEIIPSVVIVLGTASEGKVN- VAGVTKDLINK 828

Query: 828 --HAGNMIGLAPIVAGRGGKPDMAAGGSDASKIAELLAABAE 870  
 HAG ++K +A G GGG+PDMA AGG K+ + L+ V E  
 Sbjct: 829 GYHAGKLVKEVATRCGGGGGPRPDMAQAGGKQPEKLQDALSFVYE 873

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4505> which encodes the amino acid sequence <SEQ ID 4506>. Analysis of this protein sequence reveals the following:

55 Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.41 Transmembrane 805 - 821 ( 804 - 822)  
 ----- Final Results -----  
 60 bacterial membrane --- Certainty=0.2763 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

65 Identities = 862/870 (99%), Positives = 864/870 (99%)

-1619-

Query: 1 MKELSSAQIRQMWLD FWKSKGHSVEPSANLVPVNDPTLLWINSVATLKKYFDGGSVIPEN 60  
 MKELSSAQIRQMWLD FWKSKGH VEPSANLVPVNDPTLLWINSVATLKKYFDGGSVIPEN  
 Sbjct: 1 MKELSSAQIRQMWLD FWKSKGHCVEPSANLVPVNDPTLLWINSVATLKKYFDGGSVIPEN 60

5 Query: 61 PRITNAQKSIRTNNDIENVGKTARHHTMFEMLGNF SIGDYFRDEAIEWGFELLTSPEWFD 120  
 PRITNAQKSIRTNNDIENVGKTARHHTMFEMLGNF SIGDYFRDEAIEWGFELLTSPEWFD  
 Sbjct: 61 PRITNAQKSIRTNNDIENVGKTARHHTMFEMLGNF SIGDYFRDEAIEWGFELLTSPEWFD 120

10 Query: 121 PKDKLYMTYYPD DKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180  
 PKDKLYMTYYPD DKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD  
 Sbjct: 121 PKDKLYMTYYPD DKDSYNRWIACGVEPSHLVPIEDNFWEIGAGPSGPDTEIFFDRGEDFD 180

15 Query: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLA 240  
 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLA  
 Sbjct: 181 PENIGLRLLAEDIENDRYIEIWNIVLSQFNADPAVPRSEYKELPNKNIDTGAGLERLA 240

20 Query: 241 MQGAKTNFETDLFMP IIREVEKLSGKTYDPDGNMSFKVIADHIRALSFAIGDGALPGNE 300  
 MQGAKTNFETDLFMP IIREVEKLSGKTYDPDGNMSFKVIADHIRALSFAIGDGALPGNE  
 Sbjct: 241 MQGAKTNFETDLFMP IIREVEKLSGKTYDPDGNMSFKVIADHIRALSFAIGDGALPGNE 300

25 Query: 301 GRGYVLRRLRRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE 360  
 GRGYVLRRLRRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE  
 Sbjct: 301 GRGYVLRRLRRRAVMHGRRLGINETFLYKLVPTVGQIMESYYPEVLEKRD FIEKIVKREE 360

30 Query: 361 ETFARTIDAGSGHLD SLLAQLKAEGKDTLEGKDIFKLYDTYGF PVELTEELAEDAGYKID 420  
 ETFARTIDAGSGHLD SLLAQLKAEGKDTLEGKDIFKLYDTYGF PVELTEELAEDAGYKID  
 Sbjct: 361 ETFARTIDAGSGHLD SLLAQLKAEGKDTLEGKDIFKLYDTYGF PVELTEELAEDAGYKID 420

35 Query: 421 HEGFKSAMKEQQDRARA AVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSV IADNE 480  
 HEGFKSAMKEQQDRARA AVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSV IADNE  
 Sbjct: 421 HEGFKSAMKEQQDRARA AVVKGGSMGMQNETLAGIVEESRFEYDTYSLESSLSV IADNE 480

40 Query: 481 RTEAVSEGQALLVFAQT PPFYAEMGGQVADHGVIKNDKGDTVAEVVDVQKAPNGQPLHTVN 540  
 RTEAVSEGQALLVFAQT PPFYAEMGGQVAD G IKNDKGDTVAEVVDVQKAPNGQPLHTVN  
 Sbjct: 481 RTEAVSEGQALLVFAQT PPFYAEMGGQVADTGRIKNDKGDTVAEVVDVQKAPNGQPLHTVN 540

45 Query: 541 VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHN VIGEHAHQAGSLNEEEFLRFD 600  
 VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHN VIGEHAHQAGSLNEEEFLRFD  
 Sbjct: 541 VLASLSVGTNYTLEINKERRLAVEKNHTATHLLHAALHN VIGEHAHQAGSLNEEEFLRFD 600

50 Query: 601 FTHFEAVSNEELRHIEQEVNEQIWNLTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ 660  
 FTHFEAVSNEELRHIEQEVNEQIWN LTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ  
 Sbjct: 601 FTHFEAVSNEELRHIEQEVNEQIWNLTITTTETDVETAKEMGAMALFGEKYGKVVRVVQ 660

55 Query: 661 IGNYSVELCGGTHLNNSS EIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI 720  
 IGNYSVELCGGTHLNNSS EIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI  
 Sbjct: 661 IGNYSVELCGGTHLNNSS EIGLFKIVKEEGIGSGTRRIIAVTGRQAFEAYRNQEDALKEI 720

60 Query: 721 AATVKAPQLKDA AAKVQALSDSLRDLQKENVELKEKAAAAAAGDVFKDIQEAKGVRFIAS 780  
 AATVKAPQLKDA AAKVQALSDSLRDLQKEN ELKEKAAAAAAGDVFKD+QEAKGVRFIAS  
 Sbjct: 721 AATVKAPQLKDA AAKVQALSDSLRDLQKENAELKEKAAAAAAGDVFKDVQEAKGVRFIAS 780

Query: 781 QVDVADAGALRTFADNWKQKDYSDV LVLVAAIGEKVNVLVASKTKDVHAGNMIKGLAPIV 840  
 QVDVADAGALRTFADNWKQKDYSDV LVLVAAIGEKVNVLVASKTKDVHAGNMIK LAPIV  
 Sbjct: 781 QVDVADAGALRTFADNWKQKDYSDV LVLVAAIGEKVNVLVASKTKDVHAGNMIKELAPIV 840

Query: 841 AGRGGGKPDMA MAGGSDASKIAELLA AAVAE 870  
 AGRGGGKPDMA MAGGSDASKIAELLA AAVAE  
 Sbjct: 841 AGRGGGKPDMA MAGGSDASKIAELLA AAVAE 870

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1620-

**Example 1467**

A DNA sequence (GBSx1553) was identified in *S.agalactiae* <SEQ ID 4507> which encodes the amino acid sequence <SEQ ID 4508>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2974 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9747> which encodes amino acid sequence <SEQ ID 9748> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15920 GB:Z99123 yxjI [Bacillus subtilis]

Identities = 42/144 (29%), Positives = 73/144 (50%), Gaps = 2/144 (1%)

Query: 17 IKEKMFSLGGKFTITDLTGLPCYHVEGSLFPLPKTFKVFDEEEHLISQIEKKVLSFLPKF 76

+K+KMFS F I D + VEG F L + ++ D + IE+K++S LP++

Sbjct: 6 MKQKMFSFKDAFHIYDRDEQETFKVEGRFFSLGDSLQMTDSSGKTLVSIEQKLMSLLPRY 65

Query: 77 NVTLANGNHFTIKKDFSFLKPHYTIEDLDMEVKGNFWDMDFQLLKDNQVIANISQQWFRM 136

+++ + K +F KP + I L+ E+ G+ W +FQL V ++S++W

Sbjct: 66 EISIGGKTVCEVTKKVTFSPKPFVISGLNWEIDGDLWRDEFQLTDGENVRMSVSKKWLWSW 125

Query: 137 TSTYQVEVYSETYNDLTISLVIAI 160

+Y +++ E D+ I IAI

Sbjct: 126 GDSYHLQIAYE--EDVLICTAIAI 147

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1468**

A DNA sequence (GBSx1554) was identified in *S.agalactiae* <SEQ ID 4509> which encodes the amino acid sequence <SEQ ID 4510>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3833 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA36674 GB:AB016282 ORF17 [bacteriophage phi-105]

Identities = 45/133 (33%), Positives = 74/133 (54%), Gaps = 5/133 (3%)

Query: 2 KYTYLALFEVDKENGGINISFPDFHGAFSEADSLNEAIFNAREVLEIYTIMFEDEGKEFP 61

+Y Y ALF+ D + G ++FPD G + +S EA+ A+E + ++ FE +G P

Sbjct: 5 RYIYPALFDYDDD--GITVTFFDLPGCITFGNSGGEALTMKEAMALHLYGFEQDGDIIIP 62

Query: 62 KASSFKALASNLASDEDVIQAI SVDTELVRERERSKIVNKT VTLPSWLVEVGKENKVNFS 121

+A+ K + A + + I R + V KT+T+P W+ ++ KE+KVN+S

Sbjct: 63 EATPSKEIK---AEESQSVLLIETWMPFPRHDMENAAVKKTLTIPRWMDIAKEHKVNYS 119

-1621-

Query: 122 QLLQKAIREELQV 134  
 QLLQ+AI+E L +  
 Sbjct: 120 QLLQEAIKEHLGI 132

5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1469

10 A DNA sequence (GBSx1555) was identified in *S.agalactiae* <SEQ ID 4511> which encodes the amino acid sequence <SEQ ID 4512>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1484 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:BAA25696 GB:AB010712 NADH oxidase/alkyl hydroperoxidase  
 reductase [Streptococcus mutans]  
 Identities = 383/509 (75%), Positives = 441/509 (86%)

25 Query: 1 MVL DKEI KAQLAQYLDLLES DIVLQADLGDNDNSQKV KDFLDEIVAMSDRISLESTHLKR 60  
 M LD EIK QL QYL LLES+IVLQA L D+ NSQKV K+FL EIVAMS ISLE L R  
 Sbjct: 1 MALDAEIKEQLGQYLQLLESEIVLQAQLKDDANSQKVKEFLQEIVAMSPMISLEEKELPR 60

30 Query: 61 QPSFGIAKKGHESRVIFSGLPMGHEFTSFILALLQVSGRAPKVD EDI IKRIKGIEKTINL 120  
 PSF IAKKG ES V F+GLP+GHEFTSFILALLQVSGR PKV+ DI+KRI+ +++ ++  
 Sbjct: 61 TPSFR IAKKGQESGVEFAGLPLGHEFTSFILALLQVSGRPPKVETDIVKRIQAVDEPMHF 120

35 Query: 121 ETYVSLTCHNCPDVVQAFNIMAVLNPNIHTMIEGGM YQDEVKSKGIMSVPTVYKDQEEF 180  
 ETYVSLTCHNCPDVVQAFNIM+V+NPNI+HTM+EGGM++DE+++KGIMSVPTVYKD EF  
 Sbjct: 121 ETYVSLTCHNCPDVVQAFNIMSVVNPNIHTMVEGGMFKDEIEAKGIMSVPTVYKDGTEF 180

40 Query: 181 TSGRATIEQLLEQLDGLD AEFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGILAET 240  
 TSGRA+IEQLL+ + GPL +AF DKG V+DVLVIGGGPAGNSAAIYAARKG+KTG+LAET  
 Sbjct: 181 TSGRASIEQLLDLIAGPLKEDAFDDKGVFDVLVIGGGPAGNSAAIYAARKGVKTGLLAET 240

45 Query: 241 FGGQVIETVGIENMIGTLYTEGPKLMAQIEHTKSYDIDI IKSQLATGIEKKELVEVTLA 300  
 GGQV+ETVGIENMIGT Y EGP+LMAQ+EEHTKSY +DI+K+ A I+K +LVEV L  
 Sbjct: 241 MGGQVME TVGIENMIGTPYVEGPQLMAQVEHTKSYSDIMKAPRAKSIQKTDLVEVELD 300

50 Query: 301 NGAILQAKTAILALGAKWRNINVPGE EEFNRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGM 360  
 NGA L+AKTA+LALGAKWR INVPEGE+EF NKGVTYCPHCDGPLF K VAVIGGGNSG+  
 Sbjct: 301 NGAHLKAKTAVLALGAKWRKINVPGEKEFFNKGVTYCPHCDGPLFTDKKVAVIGGGNSGL 360

55 Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNL TILKNVATKDIVGEDHVTGLNY 420  
 EAA+DLAG+ HV +LEFLPELKAD++LQ+RA DN+TIL NVATK+I+G DHV GL Y  
 Sbjct: 361 EAAIDLAGLASHVYIILEFLPELKADKILQDRAEALDNITILT NVATKEIIGNDHVEGLRY 420

Query: 421 TDRDTNEEKHIDLEGVVFQIGLVPSTSWLKDSGIELNERQEIVVDKFGSTNIPGIFAAGD 480  
 +DR TNEE +DLEGVVFQIGLVPST WLKDSG+ LNE+ EI+V K G+TNIP IFAAGD  
 Sbjct: 421 SDRITTNEEYLLDLEGVVFQIGLVPSTDWLKDSLALNEKGEIIVAKDGATNIPAI FAAGD 480

Query: 481 CTDAAYKQIIISMGS GATAAIGAFDYLR 509  
 CTD+AYKQIIISMGS GATAA+GAFDYLR  
 Sbjct: 481 CTDSAYKQIIISMGS GATAALGAFDYLR 509

-1622-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4513> which encodes the amino acid sequence <SEQ ID 4514>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0654(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 419/510 (82%), Positives = 472/510 (92%)

```

Query: 1  MVLDEKEIKAQLAQYLDLLESDIVLQADLGDNDNSQKVKDFLDEIVAMSDRISLESTHLKR 60
          M L  +IK QLAQYL LLE+D+VLQ LGDN+ SQKVKDF++EI AMS+RIS+E+  L R
Sbjct: 1  MALSPDIKEQLAQYLTLLLEADLVLQVSLGDNEQSQKVKDFVEEIAAMSERISIENTILDR 60

Query: 61  QPSFGIAKKGHESRVIFSGPLPMGHEFTSFILALLQVSGRAPKVDDEDIIRIKRIKIEKTINL 120
          QPSF +AKKGH S V+F+GLP+GHE TSFILALLQVSGRAPKVD+D+I RIK I++ ++
Sbjct: 61  QPSFKVAKKGHSGSVFAGLPLGHELTFSFILALLQVSGRAPKVDQDVIDRIKAIDRPLHF 120

Query: 121 ETYVSLTCHNCPDVVQAFNIMAVLNPNITHMTIEGGMYQDEVKSKGIMSVPTVYKDQEEF 180
          ETYVSLTCHNCPDVVQA NIM+VLN I+HTM+EGGM+QDEVK+KGIMSVPTV+ D EEF
Sbjct: 121 ETYVSLTCHNCPDVVQALNIMSVLNDKISHTMVEGGMFQDEVKAKGIMSVPTVFLDGEEF 180

Query: 181 TSGRATIEQLLEQLDGPLDAEAFADKGVYDVLVIGGGPAGNSAAIYAARKGLKTGILAET 240
          TSGRATIEQLLEQ+ GPL EAFADKG+YDVLVIGGGPAGNSAAIYAARKGLKTG+LAET
Sbjct: 181 TSGRATIEQLLEQIAGPLSEEAFAADKGLYDVLVIGGGPAGNSAAIYAARKGLKTGLLAET 240

Query: 241 FGGQVIETVGIENMIGTLYTEGPKLMAQIEEHTKSYDIDIKSQLATGIEKKELVEVTLA 300
          FGGQV+ETVGIENMIGTLYTEGPKLMA++E HTKSYD+DIIK+QLAT IEKKE +EVTLA
Sbjct: 241 FGGQVMETVGIENMIGTLYTEGPKLMAEVEAHTKSYDVDIKAQLATSIEKKENIEVTLA 300

Query: 301 NGAILQAKTAILALGAKWRNINVPGEEEFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGM 360
          NGA+LQAKTAILALGAKWRNINVPGE+EFRNKGVTYCPHCDGPLFEGKDVAVIGGGNSG+
Sbjct: 301 NGAVLQAKTAILALGAKWRNINVPGEDEFNRNKGVTYCPHCDGPLFEGKDVAVIGGGNSGL 360

Query: 361 EAALDLAGVTKHVTVLEFLPELKADQVLQERAAKTDNLTILKNVATKDIVGEDHVTGLNY 420
          EAALDLAG+ KHV VLEFLPELKAD+VLQ+RAAKT+N+TI+KNVATKDIVGEDHVTGLNY
Sbjct: 361 EAALDLAGLAKHVYVLEFLPELKADKVLQDRAAKTNNTI+KNVATKDIVGEDHVTGLNY 420

Query: 421 TDRDRTNEEKHIDLEGVVFQIGLVPSTSWLKDSGIELNERQEIVVDKFGSTNIPGIFAAGD 480
          T+RD+ E+KH+DLEGVVFQIGLVP+T+WLKDSG+ L +R EI+VDK GSTNIPGIFAAGD
Sbjct: 421 TERDSGEDKHLIDLEGVVFQIGLVPNTAWLKDSGVNLTDRGEIIVDKHGSTNIPGIFAAGD 480

Query: 481 CTDAAYKQIIISMMSGATAAIGAFDYLRQ 510
          CTD+AYKQIIISMMSGATAAIGAFDYLRQ
Sbjct: 481 CTDSAYKQIIISMMSGATAAIGAFDYLRQ 510

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1470

A DNA sequence (GBSx1556) was identified in *S.galactiae* <SEQ ID 4515> which encodes the amino acid sequence <SEQ ID 4516>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2906(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1623-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA25695 GB:AB010712 alkyl hydroperoxidase [Streptococcus mutans]  
 Identities = 167/186 (89%), Positives = 179/186 (95%)

Query: 1 MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKWA VFCFYPADFSFVCPTELGD LQEYET 60  
 MSLVGKE++EFSAQAYH G+F+TV NEDVKGKWA VFCFYPADFSFVCPTELGD LQEY T  
 Sbjct: 1 MSLVGKEMVEFSAQAYHQGEFVTNNEDVKGKWA VFCFYPADFSFVCPTELGD LQEYAT 60

Query: 61 LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGF DVLGQDGLAQRG 120  
 L+SL VEVYSVSTDTHFVHKAWHDDSDVVGTTITY MIGDPSH++SQGF+VLG+DGLAQRG  
 Sbjct: 61 LQSLGVEVYSVSTDTHFVHKAWHDDSDVVGTTITYTMIGDPSH VLSQGF EVLGEDGLAQRG 120

Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180  
 TFI+DPDG+IQMME+NADGIGRDASTLIDKVRAAQYIRQH GEVCPAKWKEGAETL PSL  
 Sbjct: 121 TFIIDPDGIIQMMEVNADGIGRDASTLIDKVRAAQYIRQHPGEVCPAKWKEGAETLKPSL 180

Query: 181 DLVGKI 186  
 DLVGKI  
 Sbjct: 181 DLVGKI 186

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4517> which encodes the amino acid sequence <SEQ ID 4518>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3022 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 173/186 (93%), Positives = 181/186 (97%)

Query: 1 MSLVGKEIIEFSAQAYHDGKFITVTNEDVKGKWA VFCFYPADFSFVCPTELGD LQEYET 60  
 MSL+GKEI EFSAQAYHDGKFITVTNEDVKGKWA VFCFYPADFSFVCPTELGD LQEYET  
 Sbjct: 1 MSLIGKEIAEFSAQAYHDGKFITVTNEDVKGKWA VFCFYPADFSFVCPTELGD LQEYET 60

Query: 61 LKSLDVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQGF DVLGQDGLAQRG 120  
 LKSL VEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQ F+VLG+DGLAQRG  
 Sbjct: 61 LKSLGVEVYSVSTDTHFVHKAWHDDSDVVGTTITYPMIGDPSHLISQAF EVLGEDGLAQRG 120

Query: 121 TFIIDPDGVIQMMEINADGIGRDASTLIDKVRAAQYIRQHTGEVCPAKWKEGAETLTPSL 180  
 TFI+DPDG+IQMMEINADGIGRDASTLIDK+ AAQY+R+H GEVCPAKWKEGAETLTPSL  
 Sbjct: 121 TFIIDPDGIIQMMEINADGIGRDASTLIDKIHA AQYVRKHPGEVCPAKWKEGAETLTPSL 180

Query: 181 DLVGKI 186  
 DLVGKI  
 Sbjct: 181 DLVGKI 186

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1471**

A DNA sequence (GBSx1557) was identified in *S.agalactiae* <SEQ ID 4519> which encodes the amino acid sequence <SEQ ID 4520>. This protein is predicted to be 30S ribosomal protein S2 (rpsB). Analysis of this protein sequence reveals the following:

Possible site: 60

-1624-

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10           >GP:CAA50276 GB:X70925 30S ribosomal protein [Pediococcus  
                   acidilactici]  
           Identities = 190/260 (73%), Positives = 226/260 (86%), Gaps = 4/260 (1%)

Query: 1   MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60  
           M+VISMKQLLEAGVHFGHQTRRWNPKM +IFTERNGI++IDLQ+TVKL D AY FV+D

15   Sbjct: 1   MSVISMKQLLEAGVHFGHQTRRWNPMAKPFIFTERNGIYIIDLQKTVKLIDNAYNFVKDV 60

Query: 61   AANDAVILFVGTTKKQAAEAVAEEAKRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120  
           AAND V+LFVGTTKKQA A+ EEAKRAGQ+++NHRWLGGLTNW TQKRI RLK++K+M

20   Sbjct: 61   AANDGVVLFVGTTKKQAQTAIEEAKRAGQFYVNRWLGGLTNWNTIQKRIKRLKDLKMM 120

Query: 121   EEEGTFELLPKKEVALLNKQARLEKFLGGIEDMPRIIPDVMYVVDPHKEQIAVKEAKKLG 180  
           EE+GTF+ LPKKEVALLNKQ+ +LEKFLGGIEDMP IPDV++VVDP KEQIA+KEA+KL

25   Sbjct: 121   EEDGTFDRLLPKKEVALLNKQDKLEKFLGGIEDMPHIPDVLFPVDPKEQIAIKEAQKLN 180

Query: 181   IPVVAMVDTNADPDDIDVIIPANDDAIRAVKLITSKLADAVIEGRQGEDADV----DFAQ 236  
           IPVVAMVDTN DED +DVIIP+NDDAIRAV+LITSK+ADAV+EGRQGED +    + A+

30   Sbjct: 181   IPVVAMVDNTDQVDVIIPSNDDAIRAVRLITSKMADAVVEGRQGEDDEAVQEEVAE 240

Query: 237   EAQADSIEEIVEVVEGSNND 256  
           DS+E++ + VE +N+

30   Sbjct: 241   GVSKDSLEDLKKTVVEGSNE 260

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4521> which encodes the amino acid  
 sequence <SEQ ID 4522>. Analysis of this protein sequence reveals the following:

35   Possible site: 60  
       >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40                   bacterial cytoplasm --- Certainty=0.4462(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 241/254 (94%), Positives = 248/254 (96%)

45   Query: 1   MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60  
           MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA  
       Sbjct: 1   MAVISMKQLLEAGVHFGHQTRRWNPMAKYIFTERNGIHVIDLQQTVKLADQAYEFVRDA 60

50   Query: 61   AANDAVILFVGTTKKQAAEAVAEEAKRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120  
           AANDAVILFVGTTKKQAAEAVA+EA RAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM  
       Sbjct: 61   AANDAVILFVGTTKKQAAEAVADEATRAGQYFINHRWLGGLTNWGTIQKRIARLKEIKRM 120

55   Query: 121   EEEGTFELLPKKEVALLNKQARLEKFLGGIEDMPRIIPDVMYVVDPHKEQIAVKEAKKLG 180  
           EEEGTF++LPKKEVALLNKQARLEKFLGGIEDMPRIIPDVMYVVDPHKEQIAVKEAKKLG  
       Sbjct: 121   EEEGTFDVLPPKKEVALLNKQARLEKFLGGIEDMPRIIPDVMYVVDPHKEQIAVKEAKKLG 180

Query: 181   IPVVAMVDTNADPDDIDVIIPANDDAIRAVKLITSKLADAVIEGRQGEDADVDFAEQAQA 240  
           IPVVAMVDTNADPDDID+IIPANDDAIRAVKLIT+KLADA+IEGRQGEDADV F + QA

60   Sbjct: 181   IPVVAMVDTNADPDDIDIIIPANDDAIRAVKLITAKLADAIIEGRQGEDADVAFEADTQA 240

Query: 241   DSIEEIVEVVEGSN 254  
           DSIEEIVEVVEG N

Sbjct: 241   DSIEEIVEVVEGDN 254

-1625-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1472

- 5 A DNA sequence (GBSx1558) was identified in *S.galactiae* <SEQ ID 4523> which encodes the amino acid sequence <SEQ ID 4524>. Analysis of this protein sequence reveals the following:

Possible site: 49  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.2648(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAE73435 GB:AL139077 elongation factor TS [Campylobacter jejuni]  
Identities = 169/358 (47%), Positives = 226/358 (62%), Gaps = 19/358 (5%)

20 Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGLDKAIELLREKGMKAACKADRVAEG 60  
M EITA +VKELRE +GAG+MD K AL ET+GD DKA++LLREKG+ KAACKADR+AAEG  
Sbjct: 1 MTEITAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAACKADRLAAEG 60

Query: 61 LTGVYV--DGNVAIVIEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNEEALALTMPS 118  
L V V D A V E+N+ETDFVAKNDQF+ L +T I + EE + T+ +  
25 Sbjct: 61 LVSVKVSDFTSATVSEINSETDFVAKNDQFIALT KDITTAHIQSNLSQSV EELHSSTI-N 119

Query: 119 GETLEQAFVTATATIGEKISFRRFALVEKTDEQHFQAYQHNGGRIGVITV-----VEG 171  
G E+ + ATIGE + RRFA ++ Y H GR+GV+ V  
Sbjct: 120 GVKFEEYLSQIATIGENLVRRFATLKAGANGVNGYIHTNGRVGVVIAAACDSAEVAS 179

30 Query: 172 GDDALAKQVSMHVAAMKPTVLSYTELDAQFVHDELAQLNHKIEQDNESRAMV---NKPAL 228  
L +Q+ MH+AAM+P+ LSY +LD FV +E L ++E++NE R + NKP  
Sbjct: 180 KSRDLLRQICMHIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERRRLKDPNKPEN 239

35 Query: 229 PFLKYGSKAQLTDEVIAQAEEDIKAELAAEGKPEKIWDKIVPGKMDRFMLDNTKVDQEYT 288  
++ S+ QL+D ++ +AEE IK EL A+GKPEKIWD I+PGKM+ F+ DN+++D + T  
Sbjct: 240 KIPQFASRQLSDAILKEAEEKIKEELKAQKGKPEKIWDNIIPGKMNSFIADNSQLDSKLT 299

Query: 289 LLAQVYIMDDSKTVEAYLESV-----NAKAVAFVRFEVGEIEKASNDFEAEVAATM 340  
L+ Q Y+MDD KTVE + K V F+ FEVGEIEK + DF AEVAA +  
40 Sbjct: 300 LMGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEIEKKTEDFAAEVAAQL 357

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4525> which encodes the amino acid sequence <SEQ ID 4526>. Analysis of this protein sequence reveals the following:

45 Possible site: 38  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
50 bacterial cytoplasm --- Certainty=0.3942(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 307/344 (89%), Positives = 327/344 (94%)  
55 Query: 1 MAEITAKLVKELREKSGAGVMDAKKALVETDGLDKAIELLREKGMKAACKADRVAEG 60  
MAEITAKLVKELREKSGAGVMDAKKALVETDGD+DKA+ELLREKGMKAACKADRVAEG  
Sbjct: 33 MAEITAKLVKELREKSGAGVMDAKKALVETDGDMDKAVELLREKGMKAACKADRVAEG 92

-1626-

Query: 61 LTGVVYVDGNVAAVIEVNAETDFVAKNDQFVTLVNETAKVIAEGRPSNNEALALTMPSGE 120  
 LTGVVY GNVAAV+EVNAETDFVAKN QFV LVN TAKVIAEG+P+NN+EALAL MPSGE  
 Sbjct: 93 LTGVVYHGNVAAVVEVNAETDFVAKNAQFVELVNATAKVIAEGKPANDEALALVMPSGE 152

5 Query: 121 TLEQAFVTATATTIGEKISFRRFALVEKTDEQHFQAYQHNGGRIGVITVVEGGDDALAKQV 180  
 TL +A+V ATATTIGEKISFRRFAL+EK DEQHFQAYQHNGGRIGVI+VVEGGDDALAKQV  
 Sbjct: 153 TLAEAYVNATATTIGEKISFRRFALIEKADEQHFQAYQHNGGRIGVISVVEGGDDALAKQV 212

10 Query: 181 SMHVAAMKPTVLSYTELDAQFVHDELAQLNHKIEQDNESRAMVNKPALPFLKYGSKAQLT 240  
 SMH+AAMKPTVLSYTELDAQF+ DELAQLNH IE DNESRAMV+KPALPFLKYGSKAQL+  
 Sbjct: 213 SMHIAAMKPTVLSYTELDAQFIKDELAQLNHAIELDNESRAMVDKPALPFLKYGSKAQLS 272

Query: 241 DEVIAQAEDIKAEELAAEGKPEKIWDKI+PGKMDRFLDNTKVDQYTLAQVYIMDDSK 300  
 D+VI AE DIKAEELAAEGKPEKIWDKI+PGKMDRFLDNTKVDQ YTLAQVYIMDDSK  
 15 Sbjct: 273 DDVITAAEADIKAEELAAEGKPEKIWDKI+PGKMDRFLDNTKVDQAYTLAQVYIMDDSK 332

Query: 301 TVEAYLESVNAKAVAFVRFEVGEIEKASNDFEAEVAATMAAAL 344  
 TVEAYL+SVNAKA+AF RFEVGEIEK +NDFE+EVAATMAAAL  
 20 Sbjct: 333 TVEAYLDSVNAKAIAFAFARFEVGEIEKKANDFESEVAATMAAAL 376

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1473

A DNA sequence (GBSx1559) was identified in *S.agalactiae* <SEQ ID 4527> which encodes the amino acid sequence <SEQ ID 4528>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 30 bacterial cytoplasm --- Certainty=0.1312(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1474

A DNA sequence (GBSx1560) was identified in *S.agalactiae* <SEQ ID 4529> which encodes the amino acid sequence <SEQ ID 4530>. Analysis of this protein sequence reveals the following:

Possible site: 26  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -7.86 Transmembrane 128 - 144 ( 124 - 152)  
 INTEGRAL Likelihood = -4.57 Transmembrane 35 - 51 ( 33 - 53)  
 45 INTEGRAL Likelihood = -4.04 Transmembrane 92 - 108 ( 87 - 111)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4142(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA04953 GB:AP001511 small multidrug export related protein  
 [Bacillus halodurans]

Identities = 47/137 (34%), Positives = 71/137 (51%), Gaps = 5/137 (3%)

5

10

No corresponding DNA sequence was identified in *S.pyogenes*.

Lipop: Possible site: -1   Crend: 0  
McG: Discrim Score:       3.98  
GvH: Signal Score (-7.5): -2.35  
     Possible site: 26

```
ALOM program      count: 3 value: -7.86 threshold: 0.0
```

INTEGRAL	Likelihood = -7.86	Transmembrane	128 - 144 ( 124 - 152)
INTEGRAL	Likelihood = -4.57	Transmembrane	35 - 51 ( 33 - 53)
INTEGRAL	Likelihood = -4.04	Transmembrane	92 - 108 ( 87 - 111)
PERIPHERAL	Likelihood = 12.20		109

modified ALOM score: 2.07

\*\*\* Reasoning Step: 3

----- Final Results -----

```

bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

LPXTG motif: 105-109

The protein has homology with the following sequences in the databases:

[illegible]

426            456            486            516            546            576            591            621  
EALAIGVVGNNLVPPIIFFFARKVLEWGADKPYTGKFFTWCLKKGHSGQKLEKVAGEKGL-----FIALLLVFGIPLPG  
| | : |:|:| | ::| :  
EAYLLSVLGNIPLVPFLLLFLDYLVRATKVELLIARYR-----RVVERVERRKGVVERYGYLGLTIFVAIPLPV

40            50            60            70            80            90

651 681 711 741 771 801 831 861  
 TGAWTGTLLAASLLDWEFKHSVIAVMLGVILAGCIMGTLSTIGFNLF\*KS\*GEMTVSPF\*YLPFHQFDSKIRHLT\*AKCLI  
 ||| ||| : : || : || : || :  
 TCAGWTGTLTLLAFLQLNRLKAPLFISAGVCIAGVVLLASIGIIRLL  
 110 120 130

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1628-

**Example 1475**

A DNA sequence (GBSx1561) was identified in *S.agalactiae* <SEQ ID 4531> which encodes the amino acid sequence <SEQ ID 4532>. This protein is predicted to be CtsR protein (ctsR). Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.3672(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB91548 GB:AJ249133 CtsR protein [Lactococcus lactis]
      Identities = 74/146 (50%), Positives = 103/146 (69%), Gaps = 3/146 (2%)

      Query: 4   KNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINYVIKTRFTESRGYVVESKRG 63
                  KNTSD IE Y++ LLE++ + EIKR++LA+ F VVPSQINYVIKTRFT S+G+ VESKRG
20   Sbjct: 5   KNTSDIIEAYLRQLLEEAQVIEIKRADLANQFDVPSQINYVIKTRFTASKGFDVESKRG 64

      Query: 64  GGGYIRIAKVHFSQHQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIIITEREGNLILATS 123
                  GGGYI+I K +S +H+ + + +S + D++QLLFDE+++TEREGNL+L
25   Sbjct: 65  GGGYIKIVKYQSARHEFLTALYQKVPANLSSKAHDIVQLLFDEKVLITEREGNLLLLVI 124

      Query: 124 GDDVLGEQASVIRARMLRKLQLRDLR 149
                  D G + R M++ ++ RLDR
      Sbjct: 125 TD---GAISPFTRGIMMKSIIINRLDR 147

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4533> which encodes the amino acid sequence <SEQ ID 4534>. Analysis of this protein sequence reveals the following:

```

30   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
35          bacterial cytoplasm --- Certainty=0.2514(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

40   Identities = 117/151 (77%), Positives = 131/151 (86%)

      Query: 1   MAIKNTSDNIEEYIKSLLEQSGIAEIKRSNLADTFQVVPSQINYVIKTRFTESRGYVVES 60
                  M KNTSD+IEEYIK LL +SGIAEIKRS LAD+FQVVPSQINYVIKTRFTESRGY VES
45   Sbjct: 1   MPTKNTSDSIEEYIKELLAKSGIAEIKRSM LADS FQVVPSQINYVIKTRFTESRGYEVES 60

      Query: 61  KRGGGGYIRIAKVHFSQHQHQLFGNMLSTIGERISEQVFDDLIQLLFDEEIIITEREGNLIL 120
                  KRGGGGYIRIAKVHFSQ+H L GN+++TI + ISEQVF D IQLLFDE ++TEREGN+IL
50   Sbjct: 61  KRGGGGYIRIAKVHFSQKHHLIGNLMTIETD CISEQVFTDSIQLLFDEHLLTEREGNIIL 120

      Query: 121 ATSGDDVLGEQASVIRARMLRKLQLRDLRKG 151
                  A + DDVLG S IRARML +LLQR+DRKG
      Sbjct: 121 AVASDDVLGTDGSTIRARMLYRLLQRIDRKG 151

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1476**

A DNA sequence (GBSx1562) was identified in *S.agalactiae* <SEQ ID 4535> which encodes the amino acid sequence <SEQ ID 4536>. This protein is predicted to be ClpC (clpB-1). Analysis of this protein sequence reveals the following:

```

5   Possible site: 49
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -2.34    Transmembrane    32 - 48 ( 32 - 49)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1935(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAD01783 GB:AF023422 ClpC [Lactococcus lactis]
    Identities = 401/831 (48%), Positives = 571/831 (68%), Gaps = 52/831 (6%)

Query: 4   YSIKIQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAYE---SEVAIEEY 60
      Y+ L +F A A +Y+ +ES HLL AM .S+A A S++ I+
20  Sbjct: 8   YTPFLDRIFEKAAYAHQYQYGTIESAHLAAMATTSGSIAYSILAGMNVDSDDLIDLE 67

Query: 61  EAATIALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTEHVLLAMLVNK 120
      + ++ + + R+ L SP ++++ +A +++ AE VGTEH+L A+L +
25  Sbjct: 68  DLSSHVKVRSE-----LRFSPRAEEVVTVASFLAVHNNAEAVGTEHLLYALLQVE 118

Query: 121  DLLATRIELVGFGRQDDGESVRMVDLRKALERHAGF-TKDDIKAIYELRNPKKAKSGAS 179
      D ++L+L + + +V LRK +E+ G ++ KA+ + K AK A
30  Sbjct: 119 DGFGLQLKL-----QKINIVSLRKEIEKRTGLIVPENKKAVTTPMSKRKMAKGVAE 169

Query: 180  FSDMMKPPSTAGDLADFTDRDLSQMAVDGEIEPVGIRDKIEISRMVQVLSRKTKNNPVLVGD 239
      S+ L + DL++ A G+++P+IGR+ E+ R++ +LSR+TKNNPVLVGD
35  Sbjct: 170 -----NSSTPTLDSVSSDLTEARSGKLDPMIGREAEVDRLIHILSRRTKNNPVLVGE 222

Query: 240  AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSSVAGTRFRGDFEERMNQIADIE 299
      GVGK+A+ GLAQRI NG +P L + R++ L+M +VVAGT+FRG+FE+R+ I+ ++
40  Sbjct: 223 PGVGKSAIIEGLAQRIVNGQVPIGLMNSRIMALNMATVVAGTKFRGEFEDRLTAIVEEVS 282

Query: 300  EDGHILFIDELHTIMSGSGIDSTLDAANILKPALARGTLRTVGATTQEYQKHIEKDA 359
      D +I+FIDELHTI+G+G G+DS DAANILKPALARG + VGATT EYQK+IEKD
45  Sbjct: 283 ADPDVIIIFIDELHTIIGAGGGMDSVNDAAANILKPALARGDFQMVGATTYHEYQKYEKDE 342

Query: 360  ALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419
      AL RR A++ V+EP+ ++A IL GL+ +E +H V +D+A+ +AV ++ RY+TS+ LP
50  Sbjct: 343 ALERRLARINVDEPSPDEAIAILQLREKFEDYHQVKFTDQAISAVTLSVRYMTSRKLP 402

Query: 420  DSAIDLLDEASATVQMMIKKNAPSLLT-----EVDQAILDDDMKSA----- 460
      D AIDLLDEA+A V++++K ++ E+ +A++ D+K++
55  Sbjct: 403 DKAIDLLDEAAARVKILLKTKQNVFLEKDFVKAQEELAEAVIKLDVKASRIKEKAVEK 462

Query: 461  --SKALKASYKGKKRKPIAVTEDHIMATLSRLSGIPVEKLTQADSKKYLNLKELHKKRVI 518
      K K S K +KR+ VT+ ++A S L+G+P+ ++T+++S + +NLEKELHKKRV+
60  Sbjct: 463 ISDKIYKFSIKEKRQE--VTDQAVIAVASTLTGVPITQMTKSESDRLINLEKELHKKRVV 520

Query: 519  GQDDAVTAISRARRNQSGIRTGKRPIGSGFMFLGPTGVGKTELAKALAEVLFDDDESALIR 578
      GQ++A++A+SRAIRR +SG+ +RP+GSFMFLGPTGVGKTELAKALA+ +F E +IR
65  Sbjct: 521 GQEEAISAVSRAIRRRSGVADSRPMGSGFMFLGPTGVGKTELAKALADSVFGSEDNMIR 580

Query: 579  FDMSEYMEKFAASHLNGAPPYGVGYDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLL 638
      DMSE+MEK + S L GAPPYGVGYDEGG+LTE+VRNKPYSV+L DEVEKAH D+FN++L
70  Sbjct: 581 VDMSEFMKHSRLIGAPPYGVGYDEGGQLTERVRNKPYSVLLFDEVEKAHLDFVNIML 640

Query: 639  QVLDDGVLTDSRGRKVDFSNTIIIMTSNLGATALRDDKTGVGFGAKDISHDYTAMQKRIME 698
      Q+LDDG +TD++GRKVD FNTIIIMTSNLGATALRDDKTGVGFGAK+I+ DY+AMQ RI+E
75  Sbjct: 641 QILDDGFVTDTKGRKVDFRNTIIIMTSNLGATALRDDKTGVGFGAKNITADYSAMQSRILE 700

```

-1630-

Query: 699 ELKKAYRPEFINRIDEKVVFHSLSDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHL 758  
 ELK+ YRPEF+NRIDE +VFHSL + ++VKIM K LI L ++ + +K PSA+K +  
 Sbjct: 701 ELKRHYRPEFLNRIDENIVFHSLESQEIEQIVKIMSKSLIKRLAEQDIHVKLTPSAIKLI 760

Query: 759 AEDGYDIEMGARPLRRTTIQTQVEDHLSELLANQVKEGQVIKIGVSKGKLG 809  
 AE G+D E GARPLR+ +Q +VED LSE LL+ ++K G I IG S K+K  
 Sbjct: 761 AEVGFDPEYGARPLRKALQKEVEDLLSEQLLSGEIKAGNHISIGASNKKIK 811

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4537> which encodes the amino acid sequence <SEQ ID 4538>. Analysis of this protein sequence reveals the following:

Possible site: 44

15 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.75 Transmembrane 32 - 48 ( 32 - 48)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1702(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 285-287

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 618/814 (75%), Positives = 716/814 (87%), Gaps = 1/814 (0%)

Query: 1 MSHYSIKLQEVFRLAQFQAARYESHYLESWHLLLAMVLVHDSVAGLTFAEYBSEVAIEEY 60  
 M YS K+Q++FR AQFQAAR++SH LE+WH+LLAMV V +S+A + +EY+++VAIEEY  
 Sbjct: 1 MIMYSTKMQDIFRQAQFQAARFDSHCLETWHVLLAMVAVDNSLANMILSEYDAQVAIEEY 60

30 Query: 61 EAATILALGRAPKEEITNYQFLEQSPALKKILKLAENISIVVGAEDVGTETHVLLAMLVNK 120  
 EAA ILA+G+ PKE+++ F QS L +L A+ IS + ++VG+EHVL A+L+N  
 Sbjct: 61 EAAAILAMGKTPKEQLSRVDFRPQSKTLTNLLAFAQAISQITRDQEVGSEHVLFAILLNP 120

35 Query: 121 DLLATRILELVGFRGQDDGESV-RMVDLRKALERHAGFTKDDIKAIYELRNPKKAKSGAS 179  
 D++A+R+LE+ G++ +D+G R+ DLRKA+ERHAG++K+ IKAI+ELR PKK K+ +  
 Sbjct: 121 DIMASRLLEIAGYQIKDNGNGQPRLADLRKAIERHAGYSKEMIKAIHELKPKKTKTQGT 180

40 Query: 180 FSDMMKPPSTAGDLADFTDRDLSQMAVDGEIEPVGIRDKEISRMVQVLSRKTNNPVLVGD 239  
 FSDMMKPPSTAG+L+DFTRDL++MA G +E VIGRD+E+SRM+QVLSRKTNNPVLVGD  
 Sbjct: 181 FSDMMKPPSTAGELSDFTDRDLTEMARQGLLESVIGRDQEVSRMIQVLSRKTNNPVLVGD 240

45 Query: 240 AGVGKTALAYGLAQRIANGNIPYELRDMRVLELDMMSSVAGTRFRGDFEERMNQIIDIE 299  
 AGVGKTALAYGLAQRIANG IPYEL++MRVLELDMMSSVAGTRFRGDFEERMNQII DIE  
 Sbjct: 241 AGVGKTALAYGLAQRIANGAIPYELKEMRVLELDMMSSVAGTRFRGDFEERMNQIIDIE 300

50 Query: 300 EDGHIILFIDELHTIMSGSGSIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDA 359  
 DG IILF+DELHTIMSGSGSIDSTLDAANILKPAL+RGTL VGATTQEEYQKHIEKDA  
 Sbjct: 301 ADGQIILFVDELHTIMSGSGSIDSTLDAANILKPALSRLHMGVATTQEEYQKHIEKDA 360

55 Query: 360 ALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVTISDEAVMTAVKVAHRYLTSKNLP 419  
 ALSRRFAK+L+EEPN EDAY+IL+GLK +YE +HNV+IS+EAV TAVK+AHRYLTSKNLP  
 Sbjct: 361 ALSRRFAKILIEEPNTEDAYQILMGLKLSYETYHNVSISNEAVKTAVKMAHRYLTSKNLP 420

60 Query: 420 DSAIDLLDEASATVQMMIKKNAPSLLTEVDQAILDDDMKSASKALKASYKGKKRKPIAVT 479  
 DSAIDLLDEASA VQ M+KK+AP LT +DQA+++ DMK S+ L KG+ RKP VT  
 Sbjct: 421 DSAIDLLDEASAAVQNMVKKSAPELTLPIDQALINGDMKKVSRLLAKEAKGQMRKPTPVT 480

65 Query: 480 EDHIMATLSRLSGIPVEKLTQADSKKYLNLEKELHKRVIGQDDAVTAISRARRNQSGIR 539  
 ED I+ATLS+LSGIP+EKLTQADSKKYLNLEKELHKRVIGQD AVTAISRARRNQSGIR  
 Sbjct: 481 EDDILATLSKLSGIPLEKLTQADSKKYLNLEKELHKRVIGQDAVTAISRARRNQSGIR 540

Query: 540 TGKRPIGSFMTGLPTGVGKTELAKALAEVLFDDDESALIRFDMSEYMEKFAASHLNGAPPG 599  
 TGKRPIGSFMTGLPTGVGKTELAKALAEVLFDD+ALIRFDMSEYMEKFAAS LNGAPPG  
 Sbjct: 541 TGKRPIGSFMTGLPTGVGKTELAKALAEVLFDDAALIRFDMSEYMEKFAASRLNGAPPG 600

-1631-

Query: 600 YVG YDEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDGVLTD SRGRKVD FSNT 659  
 YVG YDEGGELT+KVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDG+LTDSRGRKVD FSNT  
 Sbjct: 601 YVG YDEGGELTQKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDGILTD SRGRKVD FSNT 660

Query: 660 IIIMTSNLGATALRDDKT VGF GAKDISHDY TAMQKRIMEELKKAYRPEFINRIDEKVVFH 719  
 IIIMTSNLGATALRDDKT VGF G KDI D+ AM+KRI+EEL+K YRPEFINRIDEKVVFH  
 Sbjct: 661 IIIMTSNLGATALRDDKT VGF GVKDIHQDHQAMEKRILEELRKTYRPEFINRIDEKVVFH 720

Query: 720 SLSQDNMRREVVKIMVKPLILALKDKGMDLKFQPSALKHLAEDGYDIEMGARPLRRTIQ TQ 779  
 SL+QDNMR+VVKIMV+PLI I +KG+ LK QP ALKHL+E GYD MGARPLRRT+QT+  
 Sbjct: 721 SLTQDNMRDVVKIMVQPLITTLAEKGITLKIQPLALKHLSEVGYDEHMGARPLRRTLQTE 780

Query: 780 VEDHLSSELLANQVKEGQVIKIGVSKGKLKFDIA 813  
 +ED LSEL+L+ ++ G +KIG+S GK L F IA  
 Sbjct: 781 IEDKLSELILSRELTSGH TLKIGLSHGKLT FHIA 814

A related GBS gene <SEQ ID 8819> and protein <SEQ ID 8820> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9  
 McG: Discrim Score: -13.52  
 GvH: Signal Score (-7.5): -2.1  
 Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -2.34 threshold: 0.0  
 INTEGRAL Likelihood = -2.34 Transmembrane 32 - 48 ( 32 - 49)  
 PERIPHERAL Likelihood = 0.95 112  
 modified ALOM score: 0.97

\*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.1935(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

47.4/69.6% over 804aa

*Listeria monocytogenes*

EGAD|136761| ClpC ATPase Insert characterized  
 GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase Insert characterized

ORF00207(298 - 2727 of 3045)  
 EGAD|136761|145854(2 - 806 of 825) ClpC ATPase {*Listeria monocytogenes*}  
 GP|1314297|gb|AAC44446.1||U40604 ClpC ATPase {*Listeria monocytogenes*}  
 %Match = 33.6  
 %Identity = 47.4 %Similarity = 69.6  
 Matches = 372 Mismatches = 229 Conservative Sub.s = 174

87 117 147 177 207 237 267 297  
 SFF\*STPIIWKYVINDWRAYQ\*TSF\*\*FDSIIIR\*RDNYRT\*RKFDSGDIR\*\*RLRRASLCY\*SSYAP\*IITITIR\*KRIP

M

327 357 387 417 447 477 507 537  
 FMSHYSIKLQEVFRLAQFQAARYESHYLESWHLLAMVLVHDSVAGLTFAEYSEVAIEEYEAATILALGRAPKEEITNY  
 :: : |:|: |:|: | | : |:|: |:|: : : | || :: |: : :| :: :|  
 MFGRFTQRAQKVLALSQEEAMRLNHSNLGTEHILGLVREGEGIAA--KALYELGISSEKVQOEVEGLIGHG-EKAVTTI  
 20 30 40 50 60 70

567 597 627 657 687 717 744 774  
 QFLEQSPALKKILKLAENISIVVGAEDVGTETHVLLAMLVKNKDLLATRIKLVGFRGQDDGESV-RMVDLRKALERHAGFT  
 |: | ||::|: : : :| |||||:|: : :| |:| :| :| :: :|  
 QYT---PRAKKVIELSMDEARKLGHTYVGTETHILGLIREGEGVAARVLSNLGISLNKARQQVLQLLGGGDA-----  
 90 100 110 120 130 140

-1632-

```

      804      834      864      894      924      954      984      1014
KDDIKAIYELRNPKKAKSGASFSDMMKPPSTAGDLADFTRDLSQMAVDGEIEFPVIGRDKEISRMVQVLSRRTKNNPVLVG
      :||      :      |      ||      |||: :| : :||| || |::|||:|||||:|
5  -----TGAGRQTNTQATPTLDSL-----RDLTVIAREDNLDPVIGRSKEIQRVIEVLSRRTKNNPVLIG
      150      160      170      180      190      200

      1044      1074      1104      1134      1164      1194      1224      1254
DAGVGKKTALAYGLAQRIANGNIPYELRDMRVLELDMMSVVAGTRFRGDFEERMNQIADIEEDGHIILFIDELHTIMSG
      : |||||:| ||||:| :| || ||: |||: ||||: |||: ||: :||: ||: ||: |||||: ||: ||
10  EPGVGKTAIAEGLAQQIVRNEVPETLRGKRVMTLDMGTVVAGTKYRGEFEDRLKKVMDEIRQAGNVILFIDELHTLIGAG
      220      230      240      250      260      270      280

      1284      1314      1344      1374      1404      1434      1464      1494
SGIDSTLDAANILKPALARGTLRTVGATTQEEYQKHIEKDAALSRRFAKVLVEEPNLEDAYEILLGLKPAYEAFHNVITIS
      | : :||:|||| |||| : :||| :||:|||| || ||| : |:| :||: || ||| : ||| | | :
15  -GAEGAIDASNILKPPLARGELQCIGATTLDEYRKYIEKORALERRFQPIKVDPTVEESIQLHGLRDRYEAHHRVAIT
      300      310      320      330      340      350      360

      1524      1554      1584      1614      1644      1674      1704
DEAVMTAVKVAHYRLTSKNLPDSALDILLDEASATVQM-----MIKKNAPSLLETVDAQAILDDDMKSASKALKASY
      |||: ||:: ||:: : || |||: ||: : ||: : : || : ||| : : : |
20  DEALEAAVRLSDRYISDRFLPKAIDVIDESGSKVRLKSFTTPKNVKEMENNLSDLKKEKDAAVQGEFEKAASLRDKEQ
      380      390      400      410      420      430      440

      1725      1737      1767      1797      1827      1857      1887
KGKK---RKPIA-----VTEDHIMATLSRLSGIPVEKLTQADSKKYLNLKELHKKRVIGQDDAVTAISR
      | || :| : |||| : : : |||| || : : | ||:| ||:|||| || |:|
30  KLKKSLDKKSLEETKANWQEQGLDHSEVTEDIVAEVVASWTGIPVAKLAETETNKLNLMEKLLHERVIGQDAAVKAVSL
      460      470      480      490      500      510      520

      1917      1947      1977      2007      2037      2067      2097      2127
AIRRNQSGIRTGKRPIGSEFMLGPTGVGKTELAKALAEVLFDDDESALIRFDMSEYMEKFAASHLNGAPPVGVYDEGGEL
      |:| :||: |||||:|||||:|||||:||||| :| || :|| |||||: : | |||||:|:|:|
35  AVRRARAGLKDKPRPIGSFIFLGPTGVGKTELARALAESMFGDEDSMIRIDMSEYMEKFSTARLVGAPPVGVYEEGGQL
      540      550      560      570      580      590      600

      2157      2187      2217      2247      2277      2307      2337      2367
TEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGVLTDSRGRKVDFSENTIIIMTSNLGATALRDDKTGVFGAKDISHDY
      ||||| |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
40  TEKVRQKPYSVLLDEIEKAHPDVFNMLLQVLDDGRLTDSKGRVVDFFRNTVIIMTSNIGAQEMKQDKSMGFNVTDPLKD
      620      630      640      650      660      670      680

      2397      2427      2457      2487      2517      2547      2577      2607
TAMQKRIMEELKKAYRPEFINRIDEKVVFHSLSQDNMREVVKIMVKPLILALKDKGMDLKFQPSALKHLAEDGYDIEMGA
      ||: |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
45  KAMEHRVLQDLKQAFRPEFINRIDEKVVFHSLQEKELQIVTLLTAQLTKRLAERDIHVKLTEGAKSIAKDGYDPEYGA
      700      710      720      730      740      750      760

      2637      2667      2697      2727      2757      2787      2817      2847
RPLRRTIQTQVEDHLSELLANQVKEGVKIGVSKGKLKFDIAKS*NIPVPMGTGILI*KENVQNILDIFL*IYEK*KD
      |||:| || :||| ||| || :| | :||| |||: :
50  RPLKRAIQKEVEDMLSELLRGNIKVGDYVEIGVKDGKLEVRKKDAPKKKTTSKKVKAK
      780      790      800      810      820

```

There is also homology to SEQ ID 258.

SEQ ID 8820 (GBS26) was expressed in *E. coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 9; MW 93.3kDa), in Figure 167 (lane 16 & 17; MW 108kDa) and in Figure 239 (lane 14; MW 108kDa). It was also expressed in *E. coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 7; MW 118kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1633-

**Example 1477**

A DNA sequence (GBSx1563) was identified in *S.agalactiae* <SEQ ID 4539> which encodes the amino acid sequence <SEQ ID 4540>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4541> which encodes the amino acid sequence <SEQ ID 4542>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 178/213 (83%), Positives = 199/213 (92%)

Query: 1 MLIVLAGTIGAGKSSLAAALGQHLGTDVFYEAVDNNPVLDLYYQDPQKYAFLQLQIFFLNK 60

MLIVLAGTIGAGKSSLAAALG+HLGTDVFYEAVDNNPVLDLYYQDP+KYAFLQLQI+FLNK

Sbjct: 1 MLIVLAGTIGAGKSSLAAALGHEHLGTDVFYEAVDNNPVLDLYYQDPKQKYAFLQLQIYFLNK 60

Query: 61 RFQSIKEAYKANNVLDRSIFEDELFLTLNYKNGNVTKTELDIIYKELLANMLEELEGMPK 120

RF+SIKEAY+A+NN+LDRSIFEDELFL LNYKNGNVTKTELDIY+ELLANMLEELEGMPK

Sbjct: 61 RFKSIKEAYQADNNILDRSIFEDELFLKLNKYKNGNVTKTELDIIYQELLANMLEELEGMPK 120

Query: 121 KRPDLLVYIDVSFDKMLERIDKGRSFEQVDNPELYDYKQVHSEYPEWYENYDVSPKI 180

KRPDLL+YIDVSFDKMLERI++RGRSFEQVD NP L YY QVH EYP WYE+Y+VSPK+

Sbjct: 121 KRPDLLIYIDVSFDKMLERIERRGRSFEQVDGNPSLEQYYHQVHGEYPTWYEDYEVSPKM 180

Query: 181 RIDGNKLDVFNKPEDLQHVLDITIDSELQKLDLL 213

+IDGN LDFV+NP+DL VL ID++L++L LL

Sbjct: 181 KIDGNSLDFVQNPQDLATVLKMDITKLKELHLL 213

A related GBS gene <SEQ ID 8821> and protein <SEQ ID 8822> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 0

McG: Discrim Score: 3.94

GvH: Signal Score (-7.5): 1.42

Possible site: 17

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 0 value: 7.69 threshold: 0.0

PERIPHERAL Likelihood = 7.69 49

modified ALOM score: -2.04

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1634-

SEQ ID 4540 (GBS9) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 5; MW 52kDa) and Figure 12 (lane 2 & 3; MW 50.3kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 6; MW 27kDa) and Figure 3 (lane 2; MW 25kDa). The GBS9-GST fusion product was purified (Figure 191, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 318), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1478

A DNA sequence (GBSx1564) was identified in *S.agalactiae* <SEQ ID 4543> which encodes the amino acid sequence <SEQ ID 4544>. Analysis of this protein sequence reveals the following:

Possible site: 36  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1182(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4545> which encodes the amino acid sequence <SEQ ID 4546>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 281/323 (86%), Positives = 305/323 (93%)

Query: 3 QLNSFMIGKVEIPHRTVLAPMAGITNSAFRTIAKEFGAGLVVMMEMISEKGLLYNNEKTL 62  
+LNSSF IG VEIPHRTVLAPMAG+TNSAFRTIAKEFGAGLVVMMEMISEKGLLYNNEKTL  
Sbjct: 27 KLNSSFRIGDVEIPHRTVLAPMAGVTNSAFRTIAKEFGAGLVVMMEMISEKGLLYNNEKTL 86

Query: 63 HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQSNTKADIVDINMGCPVNVKVKNEAGAKW 122  
HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQ+NTKADIVDINMGCPVNVKVKNEAGAKW  
Sbjct: 87 HMLHIDENEHPMSIQLFGGDAEGLKRAADFIQTNTKADIVDINMGCPVNVKVKNEAGAKW 146

Query: 123 LRDPEKIYHIVKEVTSVLDIPLTVKMRTGWSNSSNAIENALAAESAGVSALAMHGRTREQ 182  
LRDP+KIYHIVKEVTSVLDIPLTVKMRTGW+DSS A+ENALAAESAGVSALAMHGRTREQ  
Sbjct: 147 LRDPDKIYHIVKEVTSVLDIPLTVKMRTGWADSSLAIVENALAAESAGVSALAMHGRTREQ 206

Query: 183 MYTGTCDHETLKGKAVKTSIPFIANGDIRTVHDAKFMIIEIGADAIMVGRGARSNPYIF 242  
MYTGTCDHETL +V+KA+T IPFI NGD+R+V DAKFMIEEIG DA+M+GR A +NPY+F  
Sbjct: 207 MYTGTCDHETLARVSKAITKIPFIGNGDVRSVQDAKFMIIEIGVDAVMIGRAAMNNPYLF 266

Query: 243 TQINHFFETG+ LPDLPF K LD+A+DHL RL+NLKGETIAVREFRGLAPHYLRGKSGAA 302  
TQINHFFETG+ LPDLPF K LD+A+DHL RL+NLKGETIAVREFRGLAPHYLRG +GAA  
Sbjct: 267 TQINHFFETGQELPDLPPFAKLDIAKHLKRLINLKGETIAVREFRGLAPHYLRGTAGAA 326

Query: 303 KIRGAVSRAETLAEVQELFAGLR 325  
K+RGAVSRAETLAEV+ +F +R  
Sbjct: 327 KVRGAVSRAETLAEEAIFETVR 349

-1635-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1479

A DNA sequence (GBSx1565) was identified in *S.agalactiae* <SEQ ID 4547> which encodes the amino acid sequence <SEQ ID 4548>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.2164(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

There is also homology to SEQ ID 3930:

15       Identities = 235/288 (81%), Positives = 259/288 (89%)

Query: 1   MDKIIKSISTSGSFRAYVLDCTETVRTAQEKHQTLSSTVALGRTLIAHQILAANQKGNS 60  
           MDKIIKSI+ SG+FRAYVLD TETV AQEKH TLSSTVALGRTLIAHQILAANQKG+S

20       Sbjct: 1   MDKIIKSIAQSGAFRAYVLDSTETVALAQEKHNTLSSTVALGRTLIAHQILAANQKGDS 60

Query: 61   KVTVKVIGDSSFGHIISVADTKGNVKGYYQNTGVDIKKTATGEVLVGPFGMNGHFVVITD 120  
           K+TVKVIIGDSSFGHIISVADTKG+VKGYYQNTGVDIKKTATGEVLVGPFGMNGHFV I D

Sbjct: 61   KITVKVIGDSSFGHIISVADTKGHVKGYYQNTGVDIKKTATGEVLVGPFGMNGHFVTTIID 120

25       Query: 121 YATGQPYTSTTPLITGEIGEDFAYYLTESEQTTPSAVGLNVLLDDEKVKVAGGFMLQVLP 180  
           Y TG PYTSTTPLITGEIGEDFAYYLTESEQTTPSA+GLNVLLD+ DKVKVAGGFML+QVLP

Sbjct: 121 YGTGNPYTSTTPLITGEIGEDFAYYLTESEQTTPSAIGLVLLDENDKVKVAGGFMLQVLP 180

Query: 181 GASDEEISRYEKRIQEMPSSISLLSENHIESLLSAIYGEDDYKRLSEDSLAFYCDCSKE 240

30       GAS+EEI+RYEKR+QEMP+IS LL S+NH+++LL AIYG++ YKRLSE+ L+F CDCS+E

Sbjct: 181 GASEEEIARYEKRLQEMPAISHLLASKNHVDALLEAIYGDEPYKRLSEEPISFQCDCSRE 240

Query: 241 RFEAALLTLGPKELQAMKDEDKGVEITCQFCNQTYFTEEDLEKIIND 288

RFEAAL+TL +LQAM DEDKG EI CQFC Y F E DLE II+D

35       Sbjct: 241 RFEAALMTLPKADLQAMIDEDKGAEIVCQFCGKYQFNESDLEAIISD 288

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1480

40   A DNA sequence (GBSx1566) was identified in *S.agalactiae* <SEQ ID 4549> which encodes the amino acid sequence <SEQ ID 4550>. This protein is predicted to be surface-located membrane protein 1 (Imp1). Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45                   bacterial cytoplasm --- Certainty=0.4312(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB93480 GB:AF019377 tellurite resistance protein [Rhodobacter  
                   sphaeroides]

Identities = 64/350 (18%), Positives = 146/350 (41%), Gaps = 7/350 (2%)

55

-1636-

Query: 44 LTPAQKSAISEKTPALVDTFVGDQNALDFGQSAVEGVNTTVNHILSEQKKIQIPQVDDL 103  
 L A E + + V D + + + F G A + T + L + + K + D  
 Sbjct: 34 LASAPPEKAQEIRRRMAELNVSDSQSIIGFGSKAQELQTISQQMLADVKNKDVGPAGDS 93

5 Query: 104 LKNANRELNGFIAKYKDPAELEKKPNLIQKLFQSKTSLQEFYFDSQNIQKMDMMMA 163  
 L + + G F + + + K + + L + + F + + + + Q + + D +  
 Sbjct: 94 LREVVSTIRGF-----SVSEFDVRRKASWVERLLGRT-APFARFVARYEDVQQQIDRITQ 147

10 Query: 164 NVVKQEDTLARNIVSAEMLIEDNTKSIENLVGVIAFIESSQAEANRASHLQOEILALDS 223  
 + + + E L + + I + + L + L I A + A + R + + + A  
 Sbjct: 148 SLLTHEHRLKDIKGLDILYARTLDFYDELALYIAAGDEVLAIDLGRVIPAKEAEVAATP 207

15 Query: 224 QTSEYQIKSNQLARMTEVINTLEQQHPEYVSRLYVAVATTPQMRNLVKVSSDMRQKLGML 283  
 + + I K + L + + L E + + + V + P + R + + + + +  
 Sbjct: 208 E-GDRMIKAQELRDLRAARDLERRVHDLKLTRQVTMQSLPSIRLVQENDKALVTRINST 266

20 Query: 284 RRNTIPTMKLSIAQLGMMQQSVKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSP 343  
 NT+P + +AQ +Q+S + + + N L AE +A + + K +  
 Sbjct: 267 LVNTVPLWETQLAQAVTIQSRSEAAEAVRGASDLTNELLTANAENLQANKIVRKEMERG 326

Query: 344 TVSIKSVTALAESLVAQNNGIIAAIDKGRKERAQLES AVIKSAETINDSV 393  
 I + + V + L + A N + A D + GR RA E + + + + D + +  
 Sbjct: 327 VFDIEAVKKANATLIATINESLAIADGRARRATAETELQRMELRDTL 376

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4551> which encodes the amino acid sequence <SEQ ID 4552>. Analysis of this protein sequence reveals the following:

Possible site: 55  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3230(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 333/413 (80%), Positives = 379/413 (91%)

Query: 5 FNFIDIDQIADNAITKTDKTTTIIISNQTSQTGQIAFFFEKLTPAQKSAISEKTPALVDTFV 64  
 FNFIDIDQIADNA+ KTDKTT+IIS+ T GQI+FFFEKL+ Q++AI+ K PALVDTF+  
 40 Sbjct: 4 FNFIDIDQIADNAVIKTDKTTDIISDLPTDTNGQISFFFEKLSADQQTAITAKAPALVDTFL 63

Query: 65 GDQNALDLFGQSAVEGVNTTVNHILSEQKKIQIPQVDDLKKNANRELNGFIAKYKDATPA 124  
 DQNALDLFGQSAVEGVN TVNHIL+EQKK+QIPQVDDLK+ NRELNGFIAKYKDATP  
 45 Sbjct: 64 ADQNALDLFGQSAVEGVNATVNHILAEQKKLQIPQVDDLKSTNRELNGFIAKYKDATPV 123

Query: 125 ELEKKPNLIQKLFQSKTSLQEFYFDSQNIQKMDMMMAANVVKQEDTLARNIVSAEMLIE 184  
 +L+KKPN +QKLFQK+ +LQEFYFDSQNIQKMD MAA VVKQEDTLARNIVSAE+LIE  
 50 Sbjct: 124 DLDKKPNFLQKLFQSRDTLQEFYFDSQNIQKMDSMAAAVVKQEDTLARNIVSAELLIE 183

Query: 185 DNTKSIENLVGVIAFIESSQAEANRASHLQOEILALDSQTSEYQIKSNQLARMTEVINT 244  
 DNTKSIE+LVGVIAFIE+SQ EA+ RA+ LQ+++ DS T +YQIK++ LAR TEVINT  
 55 Sbjct: 184 DNTKSIEHLVGVI AFIEASQKEASQRAALQKDLTKDSATPDYQIKADLLARTTEVINT 243

Query: 245 LEQQHPEYVSRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 304  
 LEQQH EY+SRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS  
 60 Sbjct: 244 LEQQHTEYLSRLYVAVATTPQMRNLVKVSSDMRQKLGMLRRNTIPTMKLSIAQLGMMQQS 303

Query: 305 VKSGVTADAIVNANNAALQMLAETSKEAIPMLEKTAQSPTVSIKSVTALAESLVAQNNGI 364  
 VKSG+TADAI+NANNAALQMLAETSKEAIP LE++AQ+PT+S+KSVT+LAESLVAQNNGI  
 65 Sbjct: 304 VKSGMTADAIINANNAALQMLAETSKEAIPALEQSAQNPTLSMKSVTSLAESLVAQNNGI 363

Query: 365 IAAIDKGRKERAQLES AVIKSAETINDSVKIRDKKIVEALLNEGKSTQEKVDE 417  
 IAAID GRKERAQESA+I+SAETINDSVK+RD+ IV+ALL+EGK TQ+ +D+  
 70 Sbjct: 364 IAAIDHGRKERAQLES AIIRSAETINDSVKLRDQINIVQALLSEGKETQKTIDK 416

-1637-

SEQ ID 4550 (GBS201) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 5; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 3; MW 74.5kDa) and in Figure 62 (lane 8 & 9; MW 74.5kDa). The GBS201-GST fusion product was purified (Figure 209, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 304), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1481

A DNA sequence (GBSx1567) was identified in *S.agalactiae* <SEQ ID 4553> which encodes the amino acid sequence <SEQ ID 4554>. This protein is predicted to be rhopty protein. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -6.58 Transmembrane 13 - 29 ( 10 - 31)  
INTEGRAL Likelihood = -1.54 Transmembrane 33 - 49 ( 33 - 49)

----- Final Results -----

bacterial membrane --- Certainty=0.3633(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4555> which encodes the amino acid sequence <SEQ ID 4556>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 115/239 (48%), Positives = 162/239 (67%), Gaps = 3/239 (1%)

Query: 32 EVIATLLIIGGGYCAYYVVD-KKRLKRFTSNQRIEALKSDIKETDQDIRHLEILKKNRS 90  
+++ + I G GY + V +KRL + +++E LK+ I+ D+ +R L+ D+  
Sbjct: 42 DILPAIAIGGTGYAIFRVRSHQKRLAKAKIAKQLEDLKAKIQLADRKVRLLDTYLADHDD 101  
  
Query: 91 KEYIKLAHQILPQLDLIRNEANQIQKAIENPIYKRITKKANTFSNEINEQLIKLHASPEL 150  
+Y LA Q+LPQL I+ +A L+ ++P IY+RITKKAN ++I QL L + L  
Sbjct: 102 FQYNVLAQQQLLPQLSDIKAKAITLKDQLDPQIYRITKKANDVESDITLQLETIQTATTL 161  
  
Query: 151 --EPISDQEDMIRIAPELKPFYHNIQDDHFAILKKIEADNKAELAAIHQANMKRFTDV 208  
+P+ +I APELKP+Y NIQ DH AIL KI+ ADN+ EL A+H ANM+RF D+  
Sbjct: 162 NPQPLKTPSPNLINKAPELKPYYDNIQTDHQAILAKIQGADNQEELLALHDANMRRFEDI 221  
  
Query: 209 LAGYIRIKQSPKNFNNAKERLEQALQAIKKFNLDLDETFLRLNESDMKDFDVSRLMMQG 267  
L GY++IK+ PKN+ NA RLEQA QAI++F+ DLDETFLRLNESD+KDFD+SLR+MQG  
Sbjct: 222 LTGYLKIKEEPKNYYNAAARLEQAKQAIQQFDEDLDETFLRLNESDLKDFDISLRIMQG 280

-1638-

SEQ ID 4554 (GBS265) was expressed in *E.coli* as GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 2; MW 56kDa) and in Figure 62 (lane 6; MW 56.3kDa).

The GBS265-GST fusion product was purified (Figure 207, lane 5) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 258A) and FACS (Figure 258B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1482

A DNA sequence (GBSx1568) was identified in *S.agalactiae* <SEQ ID 4557> which encodes the amino acid sequence <SEQ ID 4558>. This protein is predicted to be glutamate--cysteine ligase (gshA). Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.70 Transmembrane 575 - 591 ( 575 - 591)

----- Final Results -----

bacterial membrane --- Certainty=0.1680(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG08588 GB:AE004933 glutamate--cysteine ligase [*Pseudomonas aeruginosa*]  
Identities = 142/468 (30%), Positives = 220/468 (46%), Gaps = 62/468 (13%)

Query: 12 SHLPIL-QATFGLERESLRHQPTQVAQTTPHPTKLSGRNYHPYIQTQDYSEPLELITPI 70  
++LP+L + G+ERE LR+ ++A TPHP+ LGS HP I TDYSE LE ITP  
Sbjct: 16 ANLPLLTECLHGIERECLRVDSDG-KLALTPHPRALGSTLTTHPQITTDYSEALLEFITPT 74

Query: 71 AKDSQEAIRFLKAISDVAGRSINHDEYLWPLSMPPKV-REEDIQIAQLEDA---FEYDY 125  
D + + L+ I A ++ EYLW SMP ++ EE I IA+ + +Y Y  
Sbjct: 75 ETDVADTLGDLERIHRFASSKLD-GEYLWSPSMPCELPDEESIPIARYGSSMIGRLKYVY 133

Query: 126 RKYLEKTYGKLIQSISGIHYNLGLGQELLTSLFELSQAD-NAIDFQNLQYMKLSQNFRLY 184  
RK L YGK +Q I+GIHYN L + L L + ++ + D+Q+ Y+ L +NF RY  
Sbjct: 134 RKGLALRYGKTMQCIAGIHYNFSPLERLWPLLRQAEGSELSESDYQSAAYIALIRNFRY 193

Query: 185 RWLLTYLYGASPVAEEDFLDQKLNPNVR-----SLRNSHLGYVNHKDIRIS-- 230  
WLL YL+GASP + FL + + R SLR S LGY N+ ++  
Sbjct: 194 SWLLMYLFGASPALDAGFLRGRPSQLERLDEHTLYLPYATSLRMSDLGYQNNAQAGLTPC 253

Query: 231 YTSLKDYVNDLENV-----KSGQLIAEKEFYSPVRLR-----G 264  
Y L+ Y++ L AV + L E E+YS +R + G  
Sbjct: 254 YNDLQSYIDSLRQAVSTPYPPYEKVGTKQDGEVQLNTINILQIENEYSSIRPKRVITYT 313

Query: 265 SKACRNYLEKGITYLEFRFTDLNPFSPIGITQETVDTVHLFLALLWIDS----- 314  
+ + +G+ Y+E R D+NPF P+GI + + FLL + DS  
Sbjct: 314 ERPVQALAAARGVQYVEVRCLDINPFLPLGIDLDEARFLDAFLFLCAFSDSPLINGECSDA 373

Query: 315 SSHIDQDIKEANRLN-DLIALSHPLEKLPNQAPVSDLVDMQSVIQHFNLSPYYQDLLES 373  
+ + +KE R L P+E + + + +++ + L +  
Sbjct: 374 TDNFLAVVKEGRRPGLQLQRRGQPVQLQVWANELLERIADTAALLDRAGGEAHAALAA 433

Query: 374 VKRQIQSPELTVAGQLLEMI--EGLSLETFGQRQGIYHDYAWAEPYA 419  
+ ++ ELT + Q+L+++ G S E F RQ + + +Y + P A  
Sbjct: 434 QRAKVADAELTPSAQVLKVMRERGSEFAFSLRQSREHAIFYRQHPA 481

-1639-

There is also homology to SEQ ID 4560.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1483

- 5 A DNA sequence (GBSx1569) was identified in *S.agalactiae* <SEQ ID 4561> which encodes the amino acid sequence <SEQ ID 4562>. Analysis of this protein sequence reveals the following:

Possible site: 59  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1504 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB73814 GB:AL139078 helix-turn-helix containing protein  
[Campylobacter jejuni]  
Identities = 107/223 (47%), Positives = 148/223 (65%), Gaps = 7/223 (3%)

20 Query: 1 MDKEKLDYWKTIITFLHNVLDNYEIVLHVVDENDIYIGELVNSHISGRTISSPLTTFAL 60  
MD+ + + FL VLG+ YEIV HV+ E+ YI + NSHISGR++ SPLT FA  
Sbjct: 1 MDEGQKQQFIKLTFLGEVLGEQYEIVFHVITEDGAYIAAIA NSHISGRSLDSPLTAFAS 60

25 Query: 61 DLIKNKYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120  
+L++NK Y EKDF+ +YKA+V +K +RGSTFFIKN ++L G+LCIN D S +++  
Sbjct: 61 ELMQNKKYLEKDFLCDYKALVGK-SKLIRGSTFFIKN-HDKLVGILCINHDTSIMRDLIC 118

Query: 121 DILDVLNL-NVNKILPKSPQKISLPQQEPEVEVLSGNIQDIISEIVDPSLLNQNIHLSQE 179  
++DL + ++ IL IS Q + +E LS +I+DI+ + VD S LN + LS  
30 Sbjct: 119 KMIDLEKIGDMGDIL----GNISFSQNDSSIETLSHSIEDILVQSVDSYVLSYQLSIT 174

Query: 180 VKVEIVSKLHEKGVFQKGA VSKVAEVLNISEPSVYRYLKKIE 222  
K EI KL+EKG+F +KGAV VA+ L ISEPSVYRYLKK +  
Sbjct: 175 QKEEIAEKLYEKGIFNKGAVPIVAKFLKISEPSVYRYLKKFK 217

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4563> which encodes the amino acid sequence <SEQ ID 4564>. Analysis of this protein sequence reveals the following:

Possible site: 21  
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1636 (Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/224 (75%), Positives = 198/224 (87%), Gaps = 3/224 (1%)

50 Query: 1 MDKEKLDYWKTIITFLHNVLDNYEIVLHVVDENDIYIGELVNSHISGRTISSPLTTFAL 60  
MDKE L+YWK+ITFLH+VLGDNYEI+LHV+D+NDIYIGELVNSHISGR+ SPLTTFAL  
Sbjct: 1 MDKETLNYWKTIVITFLHDVLGDNYEIIHLVIDKNDIYIGELVNSHISGRSKQSPLTTFAL 60

Query: 61 DLIKNKYKEKDFVTNYKAIVSPLNKEVRGSTFFIKNAQNELEGMLCINLDISAYQNIAL 120  
DLI NKVYKEKDFVTNYKAIVSP +KEVRGSTFFIK+ + LEGMLCINLDISAYQ +A  
55 Sbjct: 61 DLITNKVYKEKDFVTNYKAIVSPQHKEVRGSTFFIKDKKGNLEGMLCINLDISAYQGVAR 120

Query: 121 DILDVLNLNVNKILP--KSPQKISLPQQEPEVEVLSGNIQDIISEIVDPSLLNQNIHLSQ 178  
D+L LVNLN+ +P K P+ ++ PQ EE VE+L+ NIQDII +I+DPSLL N+HLSQ

-1640-

Sbjct: 121 DLLKLVNINLEHFIPTAKEPKTVT-PQPEEAVEILTSNIQDIIGQIIDPSLLRHNHLSQ 179

Query: 179 EVKVEIVSKLHEKGVFQLKGAVSKVAEVLNISEPSVYRYLKKIE 222

+VK++IV+KL+EKGVFQLKGAVSKVA++L ISEPSVYRYLKKIE

Sbjct: 180 DVKIDIVAKLYEKGVFQLKGAVSKVADILCISEPSVYRYLKKIE 223

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1484

10 A DNA sequence (GBSx1570) was identified in *S.agalactiae* <SEQ ID 4565> which encodes the amino acid sequence <SEQ ID 4566>. This protein is predicted to be regulatory protein pfoR. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

15	INTEGRAL	Likelihood = -7.80	Transmembrane	299 - 315 ( 296 - 325)
	INTEGRAL	Likelihood = -7.54	Transmembrane	172 - 188 ( 169 - 193)
	INTEGRAL	Likelihood = -7.17	Transmembrane	71 - 87 ( 66 - 98)
	INTEGRAL	Likelihood = -4.99	Transmembrane	261 - 277 ( 260 - 278)
	INTEGRAL	Likelihood = -2.81	Transmembrane	128 - 144 ( 127 - 149)
20	INTEGRAL	Likelihood = -2.18	Transmembrane	101 - 117 ( 101 - 119)
	INTEGRAL	Likelihood = -0.53	Transmembrane	198 - 214 ( 197 - 214)

----- Final Results -----

25	bacterial membrane	--- Certainty=0.4121(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]

30 Identities = 96/147 (65%), Positives = 122/147 (82%)

Query: 100 GTGIIPGFLAGYLVGFLVKWMERNIPGGDLISIIIIIGAPLTRLVAKLLTPLINSTLLTI 159  
G GI+PGF+AGYL F++K++E+ IP GLDLI II++GAPL R +A + PL+ +TL I

Sbjct: 1 GFGILPGFIAGYLGSEFVIKFKIPAGLDLIVIIIVLGAPLVRGIAAISNPLVETTLQNI 60

35 Query: 160 GDILTSGAHSNPILMGIIILGGTIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFM 219  
G ++T+ + ++PI+MGIIILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM

Sbjct: 61 GGVITATSTASPIIMMGIIILGGIVTVVATAPLSSMALTAMLGLTGLPMAIGALAVFGSSFM 120

40 Query: 220 NGVLFHKLKLGSRKDNIAFAVEPLTQA 246

N V F K+K GS+KD IA A+EPLTQA

Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4567> which encodes the amino acid  
45 sequence <SEQ ID 4568>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have a cleavable N-term signal seq.

50	INTEGRAL	Likelihood = -8.70	Transmembrane	303 - 319 ( 296 - 325)
	INTEGRAL	Likelihood = -7.11	Transmembrane	70 - 86 ( 66 - 98)
	INTEGRAL	Likelihood = -6.53	Transmembrane	172 - 188 ( 169 - 193)
	INTEGRAL	Likelihood = -4.83	Transmembrane	261 - 277 ( 260 - 278)
	INTEGRAL	Likelihood = -2.55	Transmembrane	101 - 117 ( 101 - 119)
	INTEGRAL	Likelihood = -2.28	Transmembrane	124 - 140 ( 124 - 140)
55	INTEGRAL	Likelihood = -1.91	Transmembrane	198 - 214 ( 197 - 215)

----- Final Results -----

bacterial membrane	--- Certainty=0.4482(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

-1641-

The protein has homology with the following sequences in the databases:

>GP:CAA60239 GB:X86525 pfoS [Clostridium perfringens]  
Identities = 95/147 (64%), Positives = 123/147 (83%)

Query: 100 GTGIIPGFVAGYVVSFLIKWMEKNIPGGDLISIIIVGAPLTRFLAQLITPVINSTLLTI 159  
G GI+PGF+AGY+ SF+IK++EK IP GLDLI II++GAPL R +A + P++ +TL I  
Sbjct: 1 GFGILPGFIAGYLGSEFVIKFLKIPAGLDLIVIIIVGAPLVRGIAAISNPLVETTLQNI 60

Query: 160 GDILTSSANSNPIIMGMILGGTIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFM 219  
G ++T+++ ++PI+MG+ILGG + VVATAPLSSMALTAMLGLTG+PMAIGAL+VFGSSFM  
Sbjct: 61 GGVITATSTASPIIMGIIILGGIVTVVATAPLSSMALTAMLGLTGLPMAIGALAVFGSSFM 120

Query: 220 NGVLFYRLKLGERKDNIAFAIEPLTQA 246  
N V F ++K G +KD IA AIEPLTQA  
Sbjct: 121 NLVFFGKMKFGSKKDTIAVAIEPLTQA 147

An alignment of the GAS and GBS proteins is shown below.

Identities = 302/339 (89%), Positives = 330/339 (97%)

Query: 1 MNIIIGTSLILVLAIFTLFNYKAPYGTAMGALASAACASFLVEAFQDSFFGKVLGFQF 60  
M+IIIGTSLILVLAIF+LFNYKAP+G KAMGALASAACASFLVEAFQDSFFGKVLGFQF  
Sbjct: 1 MDIIIGTSLILVLAIFSLFNYKAPHGAKAMGALASAACASFLVEAFQDSFFGKVLGFQF 60

Query: 61 LSEVGGANGSLSGVAAAILVAIAIGVTPGYAVLIGLSVSGTGIIPGFLAGYLVGFLVKWM 120  
LSEVGGANGSLSGVAAAILVAIAIGV+PGYAVLIGLSVSGTGIIPGF+AGY+V FL+KWM  
Sbjct: 61 LSEVGGANGSLSGVAAAILVAIAIGVSPGYAVLIGLSVSGTGIIPGFVAGYVVSFLIKWM 120

Query: 121 ERNIPGGDLISIIIGAPLTRVLAKLLTPLINSTLLTIGDILTSAGHSNPILMGIILGG 180  
E+NIPGGDLISIII+GAPLTR +A+L+TP+INSTLLTIGDILTS A+SNPI+MG+ILGG  
Sbjct: 121 EKNIPGGDLISIIIVGAPLTRFLAQLITPVINSTLLTIGDILTSSANSNPIIMGMILGG 180

Query: 181 TIVVVATAPLSSMALTAMLGLTGMPMAIGALSVFGSSFMNGVLFHKLKLSRKDNIAFAV 240  
TIVVVATAPLSSMALTAMLGLTG+PMAIGALSVFGSSFMNGVLF++LKLK RKDNIAFA+  
Sbjct: 181 TIVVVATAPLSSMALTAMLGLTGIPMAIGALSVFGSSFMNGVLFYRLKLGERKDNIAFAI 240

Query: 241 EPLTQADVTSANPIPIYVTNFGGAACGILIALMKLVNDTPGTATPIAGFAVMFAYNPNI 300  
EPLTQADVTSANPIPIYVTNFGGAACG+LIALMKLVNDTPGTATPIAGFAVMFAYNP+  
Sbjct: 241 EPLTQADVTSANPIPIYVTNFGGAACGVLIAMKLVNDTPGTATPIAGFAVMFAYNPVA 300

Query: 301 KVLITALGCIIISLLAGYFGGIVFKDYKLVTKKEELQARD 339  
KVLITALGCII+SL+ GY GG VFK+Y+LVTK+ELQAR+  
Sbjct: 301 KVLITALGCIIISLIVGYIGGSVFKNYRLVTKQELQARN 339

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1485

A DNA sequence (GBSx1571) was identified in *S.agalactiae* <SEQ ID 4569> which encodes the amino acid sequence <SEQ ID 4570>. This protein is predicted to be adenylosuccinate synthetase (purA). Analysis of this protein sequence reveals the following:

Possible site: 27  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0560(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-1642-

>GP:CAB16079 GB:Z99124 adenylosuccinate synthetase [Bacillus subtilis]  
Identities = 320/427 (74%), Positives = 378/427 (87%)

5 Query: 1 MTSVVVVGTTQWGDEGKGKITDFLSADA EVIARYQGGDNAGHTIVIDNKKFKLHLIPSGIF 60  
M+SVVVVVGTTQWGDEGKGKITDFLS +AEVIARYQGG+NAGHTI D +KLHLIPSGIF  
Sbjct: 1 MSSVVVVGTTQWGDEGKGKITDFLSENA EVIARYQGGNNAGHTIKFDGITYKLHLIPSGIF 60

10 Query: 61 FKEKISVINGVNVNPKSLVKELAYLHGEVTTDNLRISDRAHVILPYHIKLDQLQEDAK 120  
+K+K VINGG+VV+PK+LV ELAYLH V+TDNLRIS+RAHVILPYH+KLD+++E+ K  
Sbjct: 61 YKDKTCVINGMNVDPKALVTELAYLHERNVSTDNLRISNRAHVILPYHLKLDEVEEERK 120

15 Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVF AERLKINLA EKNRLF EKMYDSTPL 180  
G NKIGTT KGIGPAYMDKAAR+GIRIADLLDR+ FAE+L+ NL EKNRL EKMY++  
Sbjct: 121 GANKIGTTKKGIGPAYMDKAARIGIRIADLLDRDAFAEKLERNLEEKNRLL EKMYETEGF 180

20 Query: 181 EFDDIFE EY EYEGQQIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240  
+ +DI +EY EYEGQQIK+YV DTSV+LNDALD G+RVLFEGAQGVMLDIDQGTYPFVTSS  
Sbjct: 181 KLEDILDEY EY EYEGQQIKKYVCDTSVVLNDALDEGRRVLFEGAQGVMLDIDQGTYPFVTSS 240

25 Query: 241 NPVAGGVTTIGSGVGPSKINKVVGVC KAYTSRVGDGPFPTL FDEVGDRIREIGKEYGTTT 300  
NPVAGGVTTIGSGVGP+KI VVG V KAYT+RVGDGPFPTL DE+GD+IRE+G+EYGTITT  
Sbjct: 241 NPVAGGVTTIGSGVGP TKIKHVGVSKAYTTRVGDGPFPTL KDEIGDQIREVGREYGTITT 300

30 Query: 301 GRPRRVGWFD SVVMRHSRRVSGITNLSLNSIDVLSGLD TVKICVAYDL D GKRIDYYPASL 360  
GRPRRVGWFD SVV+RH+RRVSGIT+LSLNSIDVL+G++T+KICVAY G+ I+ +PASL  
Sbjct: 301 GRPRRVGWFD SVVVRHARRVSGITDLSLNSIDVLAGIETLKICVAYRYKGEIIEFPASL 360

35 Query: 361 EQLKRCKPIYEELPGWSEEDITACRSLDDL PENARNYVRRVGE L VGVRI STFSVGP GREQT 420  
+ L C+P+YEE+PGW+EDIT +SL +LPENAR+Y+ RV +L G+ +S FSVGP R QT  
Sbjct: 361 KALAECEPVYEEMP GWTEEDITGAKSLSELPENARHYLERVSQLTGIPLSIFSVGPDRSQ T 420

Query: 421 NILESVW 427  
N+L SV+  
Sbjct: 421 NVLR SVY 427

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4571> which encodes the amino acid sequence <SEQ ID 4572>. Analysis of this protein sequence reveals the following:

Possible site: 27  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0560(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 406/430 (94%), Positives = 421/430 (97%)

50 Query: 1 MTSVVVVGTTQWGDEGKGKITDFLSADA EVIARYQGGDNAGHTIVIDNKKFKLHLIPSGIF 60  
MTSVVVVVGTTQWGDEGKGKITDFLSADA EVIARYQGGDNAGHTIVID KKF LHLIPSGIF  
Sbjct: 1 MTSVVVVGTTQWGDEGKGKITDFLSADA EVIARYQGGDNAGHTIVIDGKKFKLHLIPSGIF 60

55 Query: 61 FKEKISVINGVNVNPKSLVKELAYLHGEVTTDNLRISDRAHVILPYHIKLDQLQEDAK 120  
F +KISVINGVNVNPKSLVKELAYLH EGVTTDNLRISDRAHVILPYHI+LDQLQEDAK  
Sbjct: 61 FPQKISVINGVNVNPKSLVKELAYLHDEGVTTDNLRISDRAHVILPYHIQLDQLQEDAK 120

60 Query: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDREVF AERLKINLA EKNRLF EKMYDSTPL 180  
GDNKIGTTIKGIGPAYMDKAARVGIRIADLLD+++FAERL+INLA EKNRLF EKMYDSTPL  
Sbjct: 121 GDNKIGTTIKGIGPAYMDKAARVGIRIADLLDKDIF AERLRINLA EKNRLF EKMYDSTPL 180

65 Query: 181 EFDDIFE EY EYEGQQIKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240  
+FD IFEEY YGQ+IKQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS  
Sbjct: 181 DFD AIFEEY YAGQEI KQYVTDTSVILNDALDAGKRVLFEGAQGVMLDIDQGTYPFVTSS 240

Query: 241 NPVAGGVTTIGSGVGPSKINKVVGVC KAYTSRVGDGPFPTL FDEVGDRIREIGKEYGTTT 300

-1643-

NPVAGGVTIGSGVGP+KINKVVGVCAYTSRVGDGPFPTLFDVVG+RIRE+G EYGTIT  
 Sbjct: 241 NPVAGGVTIGSGVGPKNKINKVVGVCAYTSRVGDGPFPTLFDVVGRIREVGEHYGTIT 300

Query: 301 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPA+L 360  
 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPA+L  
 Sbjct: 301 GRPRRVGWFDSVVMRHSRRVSGITNLSLNSIDVLSGLDVTVKICVAYDLGKRIDYYPANL 360

Query: 361 EQLKRCKPIYEELPGWSEIDITACRSLDDLLENARNYVRRVGLVGVRISTFSGPGREGT 420  
 EQLKRCKPIYEELPGW EDIT RSLD+LLENARNYVRRVGLVGVRISTFSGPGREGT  
 Sbjct: 361 EQLKRCKPIYEELPGWQEDITGVRSLDELLENARNYVRRVGLVGVRISTFSGPGREGT 420

Query: 421 NILESVWSNI 430  
 NILESVW++I  
 Sbjct: 421 NILESVWASI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1486

A DNA sequence (GBSx1572) was identified in *S.agalactiae* <SEQ ID 4573> which encodes the amino acid sequence <SEQ ID 4574>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.29	Transmembrane	30 - 46 ( 22 - 55)
INTEGRAL	Likelihood = -2.97	Transmembrane	110 - 126 ( 109 - 126)
INTEGRAL	Likelihood = -0.11	Transmembrane	89 - 105 ( 89 - 106)

----- Final Results -----

bacterial membrane	---	Certainty=0.4715(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8823> which encodes amino acid sequence <SEQ ID 8824> was also identified. Analysis of this protein sequence reveals the following:

Lipop Possible site: -1    Crend: 10  
 SRCFLG: 0  
 McG: Length of UR: 5  
       Peak Value of UR: 3.05  
       Net Charge of CR: 0  
 McG: Discrim Score: 4.64  
 GvH: Signal Score (-7.5): -1.66  
       Possible site: 36  
 >>> Seems to have a cleavable N-term signal seq.  
 Amino Acid Composition: calculated from 37  
 ALOM program    count: 2 value: -2.97 threshold: 0.0

INTEGRAL	Likelihood = -2.97	Transmembrane	100 - 116 ( 99 - 116)
PERIPHERAL	Likelihood = 1.38		56

modified ALOM score: 1.09  
 icm1 HYPID: 7    CFP: 0.219

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane	---	Certainty=0.2190(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has no significant homology with any sequences in the GENPEPT database and no corresponding DNA sequence was identified in *S.pyogenes*.

-1644-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1487

A DNA sequence (GBSx1573) was identified in *S.agalactiae* <SEQ ID 4575> which encodes the amino acid sequence <SEQ ID 4576>. Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.0967(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

15   No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1488

20   A DNA sequence (GBSx1574) was identified in *S.agalactiae* <SEQ ID 4577> which encodes the amino acid sequence <SEQ ID 4578>. This protein is predicted to be SgaT protein (sgaT). Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

25           INTEGRAL   Likelihood = -7.80   Transmembrane   441 - 457 ( 436 - 464)  
             INTEGRAL   Likelihood = -7.64   Transmembrane   344 - 360 ( 339 - 376)  
             INTEGRAL   Likelihood = -6.58   Transmembrane   403 - 419 ( 392 - 422)  
             INTEGRAL   Likelihood = -6.48   Transmembrane   237 - 253 ( 235 - 261)  
             INTEGRAL   Likelihood = -5.79   Transmembrane   105 - 121 ( 99 - 127)  
             INTEGRAL   Likelihood = -5.52   Transmembrane   138 - 154 ( 137 - 155)  
30           INTEGRAL   Likelihood = -4.78   Transmembrane   18 - 34 ( 14 - 38)  
             INTEGRAL   Likelihood = -2.97   Transmembrane   365 - 381 ( 365 - 383)  
             INTEGRAL   Likelihood = -0.69   Transmembrane   41 - 57 ( 41 - 57)  
             INTEGRAL   Likelihood = -0.16   Transmembrane   160 - 176 ( 160 - 176)

35           ----- Final Results -----

             bacterial membrane --- Certainty=0.4121(Affirmative) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40   The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]  
Identities = 181/451 (40%), Positives = 274/451 (60%), Gaps = 25/451 (5%)

45   Query: 11   FSQNILQNPAFFVGLLVLLIGYLLKKPLHDVFAGFIKATVGYLILNVGAGGLVNTFRPIL 70  
             F   ++ N   +G++ +GY+LL+K +   + G IK +G+++L G+G L +TF+P++  
Sbjct: 30   FFNQVMTNAPLLLGIVTCLGYILLRKSVSVIIGTKTIIGFMLLQAGSGILTSTFKPVV 89

Query: 71   VALAKKFNLEAAVIDPYFGLASANAKLETMG-FISVATTALLIGFGINILLVALRKVTKV 129  
             +++ + + A+ D Y AS A ++ MG S A+L+ +NI V LR++T +  
50   Sbjct: 90   AKMSEVYGINGAISDTY---ASMMAITDRMGDAYSWVG YAVLLALALNICYVLLRRITGI 146

Query: 130   RTLFTIGHIMVQQAATISVFVLLIPQLRNGFGAWAV----GIICGLYWAVSSNMTVEAT 185  
             RT+ +TGHIM QQA I+V + + G+ W I+ LYW ++SNM + T  
Sbjct: 147   RTIMLTGHIMFQQAGLIAVTLFIF-----GYSMWTTIICTAILVSLYWGITSNMMYKPT 200

-1645-

Query: 186 QRLTGGGGFAIGHQQQFAIWFDKVPFFGKKEENLDNLKLPFTFLNIFHDTVVASATLML 245  
 Q +T G GF+IGHQQQFA W KVAPF GKKEE+++LKLP +LNIFHD +V++A +M  
 Sbjct: 201 QEVTGCGFSIGHQQQFASWIAYKVAPFLGKKEESVEDLKLPGWLNIFHDNIVSTAIVMT 260

Query: 246 VFFGGILAVLGPDIMSNVKLIGPAGFVPTKQAFFMYILQTSLTFSVYLFIILMQGVRMFVT 305  
 +FFG IL G D + + K + +YILQT +F+V +FI+ QGVRMFV  
 Sbjct: 261 IFFGAILLSFGIDTVQ-----AMAGKVHWTVYILQTGFSFAVAIFIITQGVRFVFA 311

Query: 306 ELTNAFQGISNKLPGSFPFAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVVFNPI 365  
 EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IGQLI + +LV + I  
 Sbjct: 312 ELSEAFNGISQRLIPGAVLAIDCAIYSF-APNAVVGFMWGTIGQLIAVGILVACGSSI 370

Query: 366 LIITGFVPVFFDNAAIAVYADKRGWKAAVALSFISGIIQVALGAVAVGLLGLAGGYHGN 425  
 LII GF+P+FF NA I V+A+ GGW+AA+ + + G+I++ AV L G++ + G  
 Sbjct: 371 LIIPGFIPMFFSNATIGVFANHFGGWRAALKICLVGMIEIFGCVWAVKLTGMS-AWMGM 429

Query: 426 IDFEPPWLAFGYIFKYLGIAGYVIVCLFFLA 456  
 D+ F +GIA ++ + LA  
 Sbjct: 430 ADWSILAPPMQGFSSIGIAFMAVIVIALA 460

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4579> which encodes the amino acid sequence <SEQ ID 4580>. Analysis of this protein sequence reveals the following:

Possible site: 43  
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.51	Transmembrane	441 - 457 ( 435 - 465)
INTEGRAL	Likelihood = -7.80	Transmembrane	344 - 360 ( 339 - 376)
INTEGRAL	Likelihood = -7.64	Transmembrane	238 - 254 ( 235 - 261)
INTEGRAL	Likelihood = -5.63	Transmembrane	105 - 121 ( 100 - 127)
INTEGRAL	Likelihood = -5.52	Transmembrane	138 - 154 ( 137 - 155)
INTEGRAL	Likelihood = -5.20	Transmembrane	400 - 416 ( 392 - 422)
INTEGRAL	Likelihood = -4.78	Transmembrane	18 - 34 ( 14 - 39)
INTEGRAL	Likelihood = -2.97	Transmembrane	365 - 381 ( 365 - 383)
INTEGRAL	Likelihood = -1.49	Transmembrane	160 - 176 ( 160 - 177)
INTEGRAL	Likelihood = -0.53	Transmembrane	41 - 57 ( 41 - 57)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC77150 GB:AE000491 orf, hypothetical protein [Escherichia coli]  
 Identities = 182/461 (39%), Positives = 279/461 (60%), Gaps = 25/461 (5%)

Query: 1 MEMLLAPLNWFSQNILQNPAFFVGLLVILIGYLLKKPIYEVFAGFVKATVGYLILNVGAG 60  
 ME+L F ++ N +G++ +GY+LL+K + + G +K +G+++L G+G  
 Sbjct: 20 MEILYNIFTVFFNQVMTNAPLLLGIVTCLGYILLRKSVSVIIKGTIKTIIGFMLLQAGSG 79

Query: 61 GLVTTFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMG-FISVATTALLIGFGVNIL 119  
 L +TF+P++ +++ + + A+ D Y + A ++ MG S A+L+ +NI  
 Sbjct: 80 ILTSTFKPVVAKMSEVYGINGAISDTYASMMAT---IDRMGDAYSWVGXAVLLALALNIC 136

Query: 120 LVALRKVKTKVRTLFTTGHIMVQQAATISVFVLLLIPOFQNAFGAWAV---GIICGLYWA 175  
 V LR++T +RT+ +TGHIM QQA I+V + + + W I+ LYW  
 Sbjct: 137 YVLLRRITGIRTIMLTGHIMFQQAAGLIAVTLFIF-----GYSMWTTTICTAILVSLYWG 190

Query: 176 ISSNMTVEATQRLTGGGGFAIGHQQQFAIWFDKVPFFGKKEENLDNLKLPFTFLNIFHD 235  
 I+SNM + TQ +T G GF+IGHQQQFA W KVAPF GKKEE+++LKLP +LNIFHD  
 Sbjct: 191 ITSNNMYKPTQEVTGCGFSIGHQQQFASWIAYKVAPFLGKKEESVEDLKLPGWLNIFHD 250

Query: 236 TVVASATLMLVFFGAILAVLGPDIMSDVDLIGPAGFNPAPKQAFFMYILQTSLTFSVYLFI 295  
 +V++A +M +FFGAIL G D + + K + +YILQT +F+V +FI  
 Sbjct: 251 NIVSTAIVMTIFFGAILLSFGIDTVQAM-----AGKVHWTVYILQTGFSFAVAIFI 301

-1646-

Query: 296 LMQGVRMFVSELTNAFQGISSKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITI 355  
 + QGVRMFV+EL+ AF GIS +L+PG+ A+D AA Y F + NAV+ GF +G IGQLI +  
 Sbjct: 302 ITQGVRMFVAELSEAFNGISQRLIPGAVLAIDCAIYSF-APNAVVGFMWGTIGQLIAV 360

Query: 356 ALLVIFKNPILIIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGILQVALGAVAVGL 415  
 +IV + ILII GF+P+FF NA I V+A+ GGW+AA+ + + G++++ AV L  
 Sbjct: 361 GILVACGSSILIIPGFIPMFFSNATIGVFANHFGGWRAALKICLVMGMIEIFGCVWAVKL 420

Query: 416 LGLTGGYHGNIDLVLPLPFGYLFKFLGIAGYVLVCIFLLA 456  
 G++ + G D + P F +GIA ++ + LA  
 Sbjct: 421 TGMS-AWMGMADWSILAPFMMQGFSSIGIAFMAVIIVIALA 460

An alignment of the GAS and GBS proteins is shown below.

Identities = 437/476 (91%), Positives = 457/476 (95%)

Query: 1 MENFLAPLNWFSQNILQNPAFFVGLLVIGYLLKPLHDVFAGFIKATVGYLILNVGAG 60  
 ME LAPLNWFSQNILQNPAFFVGLLVIGYLLKPLHDVFAGFIKATVGYLILNVGAG  
 Sbjct: 1 MEMLLAPLNWFSQNILQNPAFFVGLLVIGYLLKPLHDVFAGFIKATVGYLILNVGAG 60

Query: 61 GLVNTFRPILVALAKKFNLAAVIDPYFGLASANAKLETMGFISVATTALLIGFGINILL 120  
 GLV TFRPILVALAKKF L+AAVIDPYFGLA+AN KLE MGFISVATTALLIGFG+NILL  
 Sbjct: 61 GLVNTFRPILVALAKKFELKAAVIDPYFGLAAANTKLEEMGFISVATTALLIGFGVNILL 120

Query: 121 VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPLQLRNGFGAWAVGIICGLYWAVSSNM 180  
 VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPLQ +N FGAWAVGIICGLYWA+SSNM  
 Sbjct: 121 VALRKVTKVRTLFITGHIMVQQAATISVFVLLLIPLQFQNAFGAWAVGIICGLYWAISNM 180

Query: 181 TVEATQRLTGGGGFAIGHQQQFAIWFDKVAPFFGKKEENLDNLKLPFLNIFHDTVVAS 240  
 TVEATQRLTGGGGFAIGHQQQFAIWFDKVAPFFGKKEENLDNLKLPFLNIFHDTVVAS  
 Sbjct: 181 TVEATQRLTGGGGFAIGHQQQFAIWFDKVAPFFGKKEENLDNLKLPFLNIFHDTVVAS 240

Query: 241 ATLMLVFFGGILAVLGPDIMSNVKLIGPGAFVPTKQAFFMYILQTSLTFSVYLFIHQGV 300  
 ATLMLVFFG ILAVLGPDIMSV LIGPGAF P KQAFFMYILQTSLTFSVYLFIHQGV  
 Sbjct: 241 ATLMLVFFGAILAVLGPDIMSDVDLIGPGAFNPAKQAFFMYILQTSLTFSVYLFIHQGV 300

Query: 301 RMFVTELTNAFQGISNKLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLV 360  
 RMFV+ELTNAFQGIS+KLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLV+  
 Sbjct: 301 RMFVSELTNAFQGISSKLLPGSFPAVDVAASYGFGSSNAVLSGFAFGLIGQLITIALLVI 360

Query: 361 FKNPILIIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGIIQVALGAVAVGLLGLAG 420  
 FKNPILIIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGI+QVALGAVAVGLLGL G  
 Sbjct: 361 FKNPILIIITGFVPVFFDNAAIAVYADKRGGWKAVALSFISGILQVALGAVAVGLLGLTG 420

Query: 421 GYHGNIDFEFPWLAFGYIFKYLGIAGYVIVCLFFLAIPQLQFMKSKDKEAYYRGA 476  
 GYHGNID PWL FGY+FK+LGIAGYV+VC+F LAIPQLQF K+KDKEAYYRG+A  
 Sbjct: 421 GYHGNIDLVLPLPFGYLFKFLGIAGYVLVCIFLLAIPQLQFAKADKEAYYRGEA 476

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1489

A DNA sequence (GBSx1575) was identified in *S. agalactiae* <SEQ ID 4581> which encodes the amino acid sequence <SEQ ID 4582>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1225(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1647-

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG34743 GB:AE000033 similar to PTS system: EIIB [Mycoplasma pneumoniae]  
Identities = 40/89 (44%), Positives = 62/89 (68%), Gaps = 1/89 (1%)

Query: 4 VLTACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHLIHE 63  
++ ACGNGMG+SM+IK+KVE +++LG + A S+G+ KG+ + DI+++S HL E  
Sbjct: 8 IIAACGNGMGTSMLIKIKVEKIMKELGYTAKVEA-LSMGQTKGMEHSADIIISIHILTSE 66

Query: 64 LDGRTKGHLVGLDNLMDNEIKTKLQEIL 92  
+ K +VG+ NLMD+NEIK L ++L  
Sbjct: 67 FNPNAKAKIVGVNLMDENEIKQALSKVL 95

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4583> which encodes the amino acid sequence <SEQ ID 4584>. Analysis of this protein sequence reveals the following:

Possible site: 42  
>>> Seems to have no N-terminal signal sequence  
  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.0977(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 85/92 (92%), Positives = 90/92 (97%)  
Query: 1 MVKVLTAACGNGMGSSMVIKMKVENALRQLGVSNFESASCSVGEAKGLAANYDIVVASNHL 60  
MVKVLTAACGNGMGSSMVIKMKVENALRQLGV++ +SASCSVGEAKGLA+ YDIVVASNHL  
Sbjct: 1 MVKVLTAACGNGMGSSMVIKMKVENALRQLGVTDIQSASCSVGEAKGLASGYDIVVASNHL 60  
Query: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEIL 92  
IHELDGRTKGHLVGLDNLMDNEIKTKLQE+L  
Sbjct: 61 IHELDGRTKGHLVGLDNLMDNEIKTKLQEV L 92

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1490

A DNA sequence (GBSx1576) was identified in *S.agalactiae* <SEQ ID 4585> which encodes the amino acid sequence <SEQ ID 4586>. This protein is predicted to be a pentitol phosphotransferase enzyme ii, a component (ptxA). Analysis of this protein sequence reveals the following:

Possible site: 38  
>>> Seems to have no N-terminal signal sequence  
  
----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3309(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77152 GB:AE000491 putative PTS system enzyme II A component  
[Escherichia coli K12]  
Identities = 64/150 (42%), Positives = 97/150 (64%), Gaps = 2/150 (1%)

Query: 1 MNLKQAFIENDSIRLKLSDWKEAIKLSIDPLIESGAVDAEYYDAIESTEEFGPYIIL 60  
M L+ + EN SIRL+ A W+EA+K+ +D L+ + V+ YY AI++ E+FGPY+++  
Sbjct: 1 MKLRDSLAEKNSIRLQAEAEETWQEA VKIGVDLLVAADVVEPRYYQAILDGEVQFGPYFVI 60  
Query: 61 MPMGAMPHARPEAGVKRDAFSLITLTPEPVF--PDGKEVSVLLALAATSSAIHTSVAIPQ 118

-1648-

PG+AMPH RPE GVK+ FSL+TL +P+ F D V +L+ +AA + H V I Q  
 Sbjct: 61 APGLAMPHGRPEEGVKKTGFSVTLKKPLEFNHDDNDPVDILITMAAVDANTHQEVGIMQ 120

Query: 119 IIALFELENSIQRLTECQEAKEVLAMVEES 148  
 I+ LFE E + RL C+ +EVL +++ +  
 Sbjct: 121 IVNLFEDENFDRLRACRTEQEVLDLIDRT 150

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4587> which encodes the amino acid sequence <SEQ ID 4588>. Analysis of this protein sequence reveals the following:

Possible site: 42  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2287(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 113/161 (70%), Positives = 137/161 (84%)

Query: 1 MNLKQAFIENDSIRLKLSDWKEAIKLSIDPLIESGAVDAEYDAIIESTEEFGPYIYL 60  
 MNLKQAFI+N+SIRL LSA W+EA++L++ PLI+S AV + YYDAII STE++GPYY+L  
 Sbjct: 1 MNLKQAFIDNNSIRLGLSADTQEAVALAVQLIDSKAVTSAYYDAIIASTEKGOPYVYL 60

Query: 61 MPMGAMPHARPEAGVKRDAFSLITLTPEVFPDGKEVSVLLALAATSSAIHTSVAIPQII 120  
 MPMGAMPHA GV R+AF+LITLT+PV F DGKEVSVLL LAAT +IHT+VAIPQI+  
 Sbjct: 61 MPMGAMPHAEAGLVNRNFAFALITLTKEPVTFSKGKEVSVLLTLAATDPSIHTTVAIPQIV 120

Query: 121 ALFELENSIQRLTECQEAKEVLAMVEESKNSPYLEGLDLES 161  
 ALFEL+N+I+RL CQ KEVL MVEESK+SPYLEG+DL +  
 Sbjct: 121 ALFELDNAIERLVACQSPKEVLEMVEESKDSPLYEGMDLNA 161

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1491

A DNA sequence (GBSx1577) was identified in *S.agalactiae* <SEQ ID 4589> which encodes the amino acid sequence <SEQ ID 4590>. This protein is predicted to be probable hexulose-6-phosphate synthase. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1584(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC77153 GB:AE000491 probable hexulose-6-phosphate synthase  
 [Escherichia coli K12]

Identities = 108/217 (49%), Positives = 141/217 (64%), Gaps = 3/217 (1%)

Query: 5 LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPDKIIVAD 64  
 LP LQVALD+ + A + + EVD+IE GT+ + G V L++L+P KI++AD  
 Sbjct: 3 LPMLQVALDNQTMDSAYETTRLIAEEVDIIEVGTILCVGEGVRAVRDLKALYPHKIVLAD 62

Query: 65 TKCADAGGTVAKNNAVRGADWMTICICCATIPTMEAALKAIKEERGDRGEIQIELYGDWTY 124  
 K ADAG +++ ADW+T ICCA I T + AL KE GD +QIEL G WT+  
 Sbjct: 63 AKIADAGKILSRMCFEANADWVTVICCADINTAKGALDVAKEFNGD---VQIELTGYWTW 119

-1649-

Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKKLIDMGFRVSVTGGLSTDITLQL 184  
 EQAQQW DAGI Q +YH+SRDA AG WGE D+ +K+L DMGF+V+VTGGL+ + L L  
 Sbjct: 120 EQAQQWRDAGIGQVVYHRSRDAQAGVAWGEADITAIKRLSDMGFKVTVTGGLALEDLPL 179

Query: 185 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 221  
 F+G+ + FIAGR I +A P AAR FK I +WG  
 Sbjct: 180 FKGIPIHVFIAGRSIRDAAAPVEAARQFKRSIAELWG 216

- 10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4591> which encodes the amino acid sequence <SEQ ID 4592>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.1473 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 An alignment of the GAS and GBS proteins is shown below.

Identities = 206/217 (94%), Positives = 212/217 (96%)

- Query: 5 LPNLQVALDHSDLQGAIKAAVSVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPDKIIVAD 64  
 +PNLQVALDHSDLQGA+KAAV+VGHEVDVIEAGTVCLLQVGSELVEVLRSLFP+KIIIVAD  
 25 Sbjct: 4 IPNLQVALDHSDLQGAIVKAAVAVGHEVDVIEAGTVCLLQVGSELVEVLRSLFPEKIIIVAD 63
- Query: 65 TKCADAGGTVAKNNAVRGADWMTICCATIPTMEALKAKEERGDRGEIQIELYGDWTY 124  
 TKCADAGGTVAKNNA RGADWMTICCATIPTMEALKAKEERGDRGEIQIELYGDWTY  
 30 Sbjct: 64 TKCADAGGTVAKNNAKRGADWMTICCATIPTMEALKAKEERGDRGEIQIELYGDWTY 123
- Query: 125 EQAQQWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKKLIDMGFRVSVTGGLSTDITLQL 184  
 EQAQ WLDAGISQAIYHQSRDALLAGETWGEKDLNKVK LIDMGFRVSVTGGL DTL+L  
 Sbjct: 124 EQAQLWLDAGISQAIYHQSRDALLAGETWGEKDLNKVKTLDMGFRVSVTGGLDVTDLRL 183
- 35 Query: 185 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 221  
 FEGVDVFTFIAGRGITEA+DPAAAAARAFKDEIKRIWG  
 Sbjct: 184 FEGVDVFTFIAGRGITEADDPAAAAARAFKDEIKRIWG 220

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 40 vaccines or diagnostics.

### Example 1492

A DNA sequence (GBSx1578) was identified in *S.agalactiae* <SEQ ID 4593> which encodes the amino acid sequence <SEQ ID 4594>. Analysis of this protein sequence reveals the following:

- Possible site: 36  
 45 >>> Seems to have no N-terminal signal sequence
- Final Results -----  
           bacterial cytoplasm --- Certainty=0.4179 (Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 50           bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAC22686 GB:U32783 hexulose-6-phosphate isomerase, putative  
 [Haemophilus influenzae Rd]  
 55 Identities = 143/282 (50%), Positives = 199/282 (69%), Gaps = 3/282 (1%)
- Query: 5 IGIYEKATPKHFNWLERLQFAKELGFDFVELSIDESDERLARLEWSKEERLELVKAIFET 64  
 IGIYEKA PK+ W ERL AK GF+F+E+SIDES++RL+RL W+K ER+ L ++I ++  
 Sbjct: 6 IGIYEKALPKNITWQERLSLAKACGFEBFIEMSIDESNDRLSRLNWTKSERIALHQSIQS 65

-1650-

Query: 65 GVRVPTTTFSGHRRFPMGSNNPEKEARAMDMMKKCIVFAQDIGIRNIQLAGYDVYYEEKS 124  
 G+ +P++ S HRRFP GS + + ++ ++M+K I + ++GIR IQLAGYDVYYE++  
 Sbjct: 66 GITIPSMCLSAHRRFPFGSKDKIRQKSFEIMEKAIDLSVNLGIRTIQLAGYDVYYEKQD 125

Query: 125 PETRARFIKNLRQACTWAEAAQVILSIEIMDDPFMNSIEKYLAWEKEIDSPYLFVYPDTG 184  
 ET F + + A T A AQV L++EIMD PFM+SI ++ + I+SP+ VYPD G  
 Sbjct: 126 EETIKYFQEGIEFAVTLAASAQVTLAVEIMDTPFMSSISRWKKWDTIINSPWFTVYPDIG 185

Query: 185 NVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFAVIK 244  
 N+SAW+N++ E G I+A+H+KDTY VTETSKGQFRDVPFGQGCVD+ F+++KK  
 Sbjct: 186 NLSAWNNNIEEELTLGIDKISAIHLKDTYPTVETSKGQFRDVPFGQGCVDVHFFSLLKK 245

Query: 245 TNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286  
 NY G FLIEMW+E EE I +A+ ++ MEK G+  
 Sbjct: 246 LNYRGAFIEMWTEK---NEEPLLEIIQARKWIVQQMEKAGL 284

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4595> which encodes the amino acid sequence <SEQ ID 4596>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1489(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 240/286 (83%), Positives = 271/286 (93%)

Query: 1 MTRPIGIYEKATPKHFNWLERLQFAKELGFDVVELSIDESDERLARLEWSKEERLELVKA 60  
 M RPIGIYEKATPK F W ERLQFAK+LGFDVVE+S+DESD RLARLEW+KEERL+LVKA  
 Sbjct: 15 MARPIGIYEKATPKQFTWRERLQFAKDLGFDVEMSVDESARLARLEWTKEERLDLVKA 74

Query: 61 IFETGVRVPTITFSGHRRFPMGSNNPEKEARAMDMMKKCIVFAQDIGIRNIQLAGYDVYY 120  
 I+ETG+R+PTI FSGHRR+P+GSN+P EA+++ +MK+CI AQD+G+R IQLAGYDVYY  
 Sbjct: 75 IYETGIRIPTICFSGHRRYPLGSNDPAIEAKSLKLMKQCIELAQDLGVRTIQLAGYDVYY 134

Query: 121 EEKSPETRARFIKNLRQACTWAEAAQVILSIEIMDDPFMNSIEKYLAWEKEIDSPYLFVY 180  
 E+KSPETRARFIKNLRQ+C WAEAAQV+LSIEIMDDPF+NSIEKYLAWEKEIDSPYLFVY  
 Sbjct: 135 EEKSPETRARFIKNLRQSCDWAEAAQVMLSIEIMDDPFINSIEKYLAWEKEIDSPYLFVY 194

Query: 181 PDTGNVSAWHNDLWSEFYNGHRSIAALHIKDTYAVTETSKGQFRDVPFGQGCVDWEEMFA 240  
 PD GNVSAWHNDLWSEFYNGH+SIAALH+KDTYAVTETSKGQFRDVPFGQGCVDW+E+FA  
 Sbjct: 195 PDAGNVSAWHNDLWSEFYNGHKSIAALHLKDTYAVTETSKGQFRDVPFGQGCVDWQELFA 254

Query: 241 VIKKTNYNGPFLIEMWSENCETVEETRAAIKEAQDFLYPLMEKTGV 286  
 V+KKTNYNGPFLIEMWSENC+TVEET+AAIKEAQDFLYPL+EK G+  
 Sbjct: 255 VLKKTNYNGPFLIEMWSENCETVEETKAAIKEAQDFLYPLIEKAGL 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1493

A DNA sequence (GBSx1579) was identified in *S.galactiae* <SEQ ID 4597> which encodes the amino acid sequence <SEQ ID 4598>. This protein is predicted to be L-ribulose 5-phosphate 4-epimerase. Analysis of this protein sequence reveals the following:

Possible site: 23  
 >>> Seems to have no N-terminal signal sequence

-1651-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2559(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45716 GB:AF160811 L-ribulose 5-phosphate 4-epimerase

[Bacillus stearothermophilus]

Identities = 143/229 (62%), Positives = 176/229 (76%), Gaps = 2/229 (0%)

Query: 5 LQEMRERVCEANKSLPVSILVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMVVTDL 64  
 L+E+++ V EAN LP + LV FTWGNVS +DRE GL+VIKPSGV YD+LT ++MVV DL  
 Sbjct: 2 LEEKQAVLEANLQLPQYRLVITFTWGNVSGIDRERGLVVIKPSGVAYDKLTIDDMVVVDL 61

Query: 65 EGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTHADYF 124  
 GN+VEGDL PSSD PTH+ LYK +P +GGIVHSTH A WAQAG+ IP GTTHADYF  
 Sbjct: 62 TGNVVEGDLKPSSDTPTHLWLYKQFPFGIGGIVHSTHSTWATVWAQAGKIPALGTTHADYF 121

Query: 125 YGPVPCARSLSEDEVNTAYEKETGSGVIIIEFERRDLDPMAVPGIVVRNHGPFPTWGDPAQ 184  
 YG +PC R ++ +E+ AYE ETG VI E F R LDP+ +PG++V HGPF WGDPA  
 Sbjct: 122 YGEIPCTRPMTNEEIQGA YELETGKVITETF--RFLDPLQMPGVLVHGHGPFPAWGDPA 179

Query: 185 AVYHSVVLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYGQ 233  
 AV+++VVLEEVAKM T +NP +P + ++D+HYLRKHG NAYYGQ  
 Sbjct: 180 AVHNAVVLLEEVAKMAARTYMLNPNAPKISQTLDRHYLRKHGANAYYGQ 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4599> which encodes the amino acid sequence <SEQ ID 4600>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2257(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 207/234 (88%), Positives = 220/234 (93%)

Query: 1 MAKSLQEMRERVCEANKSLPVSILVKFTWGNVSEVDREAGLIVIKPSGVDYDQLTPENMV 60  
 MAK+LQEMRERVCAANKSLP H LVKFTWGNVSEV RE G IVIKPSGVDYD LTPENMV  
 Sbjct: 1 MAKNLQEMRERVCAANKSLPQHGLVKFTWGNVSEVCRELGRIIVIKPSGVDYD LTPENMV 60

Query: 61 VTDLEGNIVEGDLNPSSDLPTHVQLYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH 120  
 VTDL+GN+VEGDLNPSSDLPTHV+LYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH  
 Sbjct: 61 VTDLDGNNVEGDLNPSSDLPTHVELYKAWPEVGGIVHSTHSTEAVGWAQAGRDIPFYGTTH 120

Query: 121 ADYFYGPVPCARSLSEDEVNTAYEKETGSGVIIIEFERRDLDPMAVPGIVVRNHGPFPTWGK 180  
 ADYFYGPVPCARSL++ EV+ AYE+ETG+VI+EEF +R LDPMAVPGIVVRNHGPFPTWGK  
 Sbjct: 121 ADYFYGPVPCARSLTKAEVDGAYEQETGNVILEEFSSKRGLDPMAVPGIVVRNHGPFPTWGK 180

Query: 181 DPAQAVYHSVVLEEVAKMNRFTQINPRVEPAPKYIMDKHYLRKHGPNAYYGQK 234  
 P QAVYHSVVLEEVA+MNR TEQINPRVEPAP+YIMDKHYLRKHGPNAYYGQK  
 Sbjct: 181 TPEQAVYHSVVLEEVARMNRLTEQINPRVEPAPRYIMDKHYLRKHGPNAYYGQK 234

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1652-

**Example 1494**

A DNA sequence (GBSx1580) was identified in *S.agalactiae* <SEQ ID 4601> which encodes the amino acid sequence <SEQ ID 4602>. This protein is predicted to be transaldolase (tal). Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10149> which encodes amino acid sequence <SEQ ID 10150> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB98962 GB:U67539 transaldolase [Methanococcus jannaschii]
Identities = 124/214 (57%), Positives = 157/214 (72%)

```

```

20 Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
    MK+FLDTA+V EI++ LG+VDGVTTNPT++++EGRDF EV+ EIC+IV+GPVSAEV
Sbjct: 1 MKFFLDTANVEEIKKYAELGLVDGVTTNPTLVAKEGRDFYEVVKEICEIVEGPVSAEVIS 60

Query: 79 LTCDEMVTAREIAKWSPNVVVKIPMTTEGLAAVSQLSKEGIKTNVTLIFTVAQGLSAMK 138
    + MV EARE+AK + N+V+KIPMT++G+ AV LS EGIKTNVTL+F+ Q L A K
25 Sbjct: 61 TDAEGMVKEAREIAKLADNIVIKIPMTKDKMKAVKILSAEGIKTNVTLVFSPLQALVAAK 120

Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
    AGAT++SPFVGRL+DIG LI D+ I Y ++E+I AS+R HV AK GA
30 Sbjct: 121 AGATYVSPFVGRLEDIGHVGMKLIEDVVKIYKNYDIKTEVIVASVRHPWHVLEAAKIGAD 180

Query: 199 IATIPDKTFASLFTHPLTDKGIETFLKDWDSFKK 232
    IAT+P LF HPLTD G+E FLKDW + K
Sbjct: 181 IATMPPAVMDKLFNHPLTDIGLERFLKDWDEYLK 214

```

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4603> which encodes the amino acid sequence <SEQ ID 4604>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 An alignment of the GAS and GBS proteins is shown below.

```

Identities = 162/214 (75%), Positives = 180/214 (83%)

```

```

50 Query: 19 MKYFLDTADVSEIRRLNRLGIVDGVTTNPTIISREGRDFKEVINEICQIVDGPVSAEVTG 78
    MK+FLDTA+V+ I+ +N LG+VDGVTTNPTIISREGRDF+ VI EIC IVDGP+SAEVTG
Sbjct: 1 MKFFLDTANVAAIKAINELGVVDGVTTNPTIISREGRDFETVIKEICDIVDGPISA EVTG 60

Query: 79 LTCDEMVTAREIAKWSPNVVVKIPMTTEGLAAVSQLSKEGIKTNVTLIFTVAQGLSAMK 138
    LT D MV EAR IAKW NVVVKIPMT EGL A + LSKEGIKTNVTLIFTV+QGL AMK
55 Sbjct: 61 LTADAMVEEARSIAKWHDNVVVKIPMTTEGLKATNLSKEGIKTNVTLIFTVSQGLMAMK 120

Query: 139 AGATFISPFVGRLEDIGTDAYALIRDLRHIIDFYGFQSEIIAASIRGLAHVEGVAKCGAH 198
    AGAT+ISPF+GRLEDIGTDAY LI DLR IID Y FQ+EIIAASIR AHVE VAK GAH
Sbjct: 121 AGATYISPFVGRLEDIGTDAYQLISDLREIIDLYDFQAEIIAASIRTTAHVEAVAKLGAH 180

```

-1653-

Query: 199 IATIPDKTFASLFTHTPLTDKGIETFLKDWDSFKK 232  
 IATIPD FA + HPLT G++TF++DW SFKK  
 Sbjct: 181 IATIPDPLFAKMTQHPLTTNGLKTFMEDWASFKK 214

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1495

A DNA sequence (GBSx1581) was identified in *S.agalactiae* <SEQ ID 4605> which encodes the amino acid sequence <SEQ ID 4606>. Analysis of this protein sequence reveals the following:

10 Possible site: 22  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1263(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB14129 GB:Z99115 transcriptional regulator (LacI family)  
 [Bacillus subtilis]  
 Identities = 108/331 (32%), Positives = 188/331 (56%), Gaps = 12/331 (3%)

Query: 6 TISDIANLVGVSKATVSYLLNGNYKKMSLQTKKIRLAIKETGYQPSKIAQSLVTKNTRT 65  
 TI D+A GVS+TVS Y+NG +S + + I+ AI E Y+PSK+AQ L K ++

25 Sbjct: 10 TIKDVAECAGVSKSTVSRYPINGKIDAI SPEKVNKIKKAI AELNYPKMAQGLKIKKSKL 69

Query: 66 IGVVIADITNPFISSVMKGIHDTCCQFGYSVNFTNSDNDIDIELENLNLNQQNVSGIIL 125  
 IG V+ADITNPF + +G+ + C Q+GYS+ N+DN + E E L +L +V G+IL

30 Sbjct: 70 IGFVVADITNPFVSVAFRGVEEVCDQYGYSIMVCNTDNSPEKEREMLLKLEAHSVEGLIL 129

Query: 126 DSVDPNHSFIETLSNDRL--VMVDRQAKDIKVDTVASDNKESTQIFLEKMQEAGYHDIYF 183  
 ++ N + + ++ +++DR+ D+K+DTV +DN+ T+ L+K+ GY D+

35 Sbjct: 130 NATGENKDVLRFAEQIPTILIDRLKLPDLKLDTVTTDNRWITKEILQKVYSKGYTDVAL 189

Query: 184 VTYPIEGISTRELRYEGFKEVVS-SNPDKLIIITE-DGSTQRILDI-----IEHSEQKP 235  
 T PI IS R R ++E+ S N + L+ + E D + L E EQK

40 Sbjct: 190 FTEPISSISPRAEAAVQEMASVQNVNGLVRLHEIDVKDKQLKAEILRSFHKEMPEQKK 249

Query: 236 GFLMMNGPTLLNFMKLNQSTVSYPEDYGLGSYEDLEWMQVLTNPVSCIKQDSYGIGCLA 295  
 L +NG +L + + + + P+D G+ ++D EW +++ P ++ I Q S+ +G A

45 Sbjct: 250 AILALNGLIMLKIISCMEELGLRIPQDIGIAGFDDTEWYKLI GPGITTIAQPSHDMGRTA 309

Query: 296 AQCLIEKISQGNEPTTARLLEVINQIVIRQS 326  
 + ++++I + + +E++ ++++R+S

50 Sbjct: 310 MERVLRKRIE--GDKGAPQTIELEAKVIMRKS 338

There is also homology to SEQ ID 2366.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### 50 Example 1496

A DNA sequence (GBSx1582) was identified in *S.agalactiae* <SEQ ID 4607> which encodes the amino acid sequence <SEQ ID 4608>. Analysis of this protein sequence reveals the following:

55 Possible site: 40  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1654-

```

bacterial cytoplasm --- Certainty=0.1661(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

5 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1497

10 A DNA sequence (GBSx1583) was identified in *S.agalactiae* <SEQ ID 4609> which encodes the amino acid sequence <SEQ ID 4610>. This protein is predicted to be GLYCERATE DEHYDROGENASE. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq

```

15

```

----- Final Results -----

```

```

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

20

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB50351 GB:AJ248287 GLYCERATE DEHYDROGENASE [Pyrococcus abyssi]
Identities = 123/325 (37%), Positives = 192/325 (58%), Gaps = 8/325 (2%)

```

25

```

Query: 1 MDKKKILVTGIVPKEGLRKLMDRFDVTYSED-RPFSRDYVLEHLSEYDGWLLM-GQKGDK 58
M K ++ +T +P+ G+ L F+V ED R R+ +LE + + D + M ++ D+
Sbjct: 1 MSKPRVFITREIPEVGIEMLEKEFEVEVWEDEREIPREILLEKVKDQVDALVTMLSERIDR 60

```

30

```

Query: 59 EMIDAGENLQIISLNAVGFHDVDTAYAKEKGIIVSNSPQAVRVPTAEMTFALILAASKRL 118
E+ + L+I++ AVG+D++D A ++GI V+N+P + TA++ FAL+LA ++ L
Sbjct: 61 EVFERAPRLRIVANYAVGYDNIDVEEATKRGIVYVNTPGVLTADATLAFALLLATARHL 120

```

35

```

Query: 119 AFYDSIVRSGEW----IDPSEQRYQGLTLQGSTLGIYGMGRIGLTVANFAKAFGMTVVYN 174
D RSGEW + + + G + G T+GI G GRIG +A A+ F M ++Y
Sbjct: 121 VKGDKFTRSGEWKKRGVAWHKFWFLGYDVYVYGTIGIIGFGRIGQAIKRARGFDMRILYY 180

```

40

```

Query: 175 DVYRLPEDKEKELGVTYLEFDQLIKTADVITIHAPALPSTIHKFNKDVFAKMKNRSYLIN 234
R PE EKEL + D+L++ +D + + P T H N++ MK + LIN
Sbjct: 181 SRTRKPE-VEKELNAEFKPLDELLRESDFVVLAVPLNKETYHMINERLKM MKRTAILIN 239

```

```

Query: 235 AARGPIVSEALIEALKEGEIAGAGLDVFENEPQVSEGLRSLDNVIMSPHAGTGTIEGRR 294
ARG ++ +ALI+ALKEG IAGAGLDV+E EP +E L SLDNV+++PH G+ T R
Sbjct: 240 VARGKVIDTKALIKALKEGWIAGAGLDVYEEEPYNEELFSLDNVVLTPHIGSATFGARE 299

```

45

```

Query: 295 TLAEEAADNIIAFDQK-PQNIIVNK 318
+A+ A+N+IAF G+ P +VN+
Sbjct: 300 GMAKLVAENLIAFKRGEVPPTLVNR 324

```

There is also homology to SEQ ID 124.

50 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1655-

**Example 1498**

A DNA sequence (GBSx1585) was identified in *S.agalactiae* <SEQ ID 4611> which encodes the amino acid sequence <SEQ ID 4612>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 34
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
   bacterial cytoplasm --- Certainty=0.1898(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1499**

A DNA sequence (GBSx1586) was identified in *S.agalactiae* <SEQ ID 4613> which encodes the amino acid sequence <SEQ ID 4614>. This protein is predicted to be PTS system, galactitol specific IIC component. Analysis of this protein sequence reveals the following:

```

20  Possible site: 25
   >>> Seems to have no N-terminal signal sequence

   INTEGRAL    Likelihood = -13.27    Transmembrane    254 - 270 ( 245 - 277)
   INTEGRAL    Likelihood = -9.24     Transmembrane     77 - 93 ( 71 - 100)
25  INTEGRAL    Likelihood = -9.24     Transmembrane    367 - 383 ( 364 - 386)
   INTEGRAL    Likelihood = -8.28     Transmembrane     32 - 48 ( 26 - 54)
   INTEGRAL    Likelihood = -7.38     Transmembrane    186 - 202 ( 182 - 215)
   INTEGRAL    Likelihood = -6.26     Transmembrane    158 - 174 ( 151 - 180)
   INTEGRAL    Likelihood = -5.79     Transmembrane    279 - 295 ( 276 - 296)
   INTEGRAL    Likelihood = -1.12     Transmembrane    342 - 358 ( 342 - 359)
30  INTEGRAL    Likelihood = -0.00     Transmembrane    308 - 324 ( 308 - 324)

   ----- Final Results -----
   bacterial membrane --- Certainty=0.6307(Affirmative) < succ>
35  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8825> which encodes amino acid sequence <SEQ ID 8826> was also identified. Analysis of this protein sequence reveals the following:

```

40  Lipop: Possible site: -1    Crend: 9
   McG: Discrim Score:      8.30
   GvH: Signal Score (-7.5): 2.97
   Possible site: 58
   >>> Seems to have a cleavable N-term signal seq.
   ALOM program    count: 9 value: -13.27 threshold: 0.0
45  INTEGRAL    Likelihood = -13.27    Transmembrane    321 - 337 ( 312 - 344)
   INTEGRAL    Likelihood = -9.24     Transmembrane    144 - 160 ( 138 - 167)
   INTEGRAL    Likelihood = -9.24     Transmembrane    434 - 450 ( 431 - 453)
   INTEGRAL    Likelihood = -8.28     Transmembrane     99 - 115 ( 93 - 121)
   INTEGRAL    Likelihood = -7.38     Transmembrane    253 - 269 ( 249 - 282)
50  INTEGRAL    Likelihood = -6.26     Transmembrane    225 - 241 ( 218 - 247)
   INTEGRAL    Likelihood = -5.79     Transmembrane    346 - 362 ( 343 - 363)
   INTEGRAL    Likelihood = -1.12     Transmembrane    409 - 425 ( 409 - 426)
   INTEGRAL    Likelihood = -0.00     Transmembrane    375 - 391 ( 375 - 391)
   PERIPHERAL    Likelihood = 0.69      188
55  modified ALOM score: 3.15

```

-1656-

\*\*\* Reasoning Step: 3

----- Final Results -----

5                   bacterial membrane --- Certainty=0.6307(Affirmative) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10           >GP:BAB03909 GB:AP001507 PTS system, galactitol-specific enzyme II,  
               C component [Bacillus halodurans]  
               Identities = 92/347 (26%), Positives = 173/347 (49%), Gaps = 15/347 (4%)

15           Query: 1    MVKTTGLHLPVIDIGWQAGSLTAFSSSEIGLSFFVFGLLIELGLFLLGITRVFVPSNLWNN 60  
                   MV   G+ L ++D+GW A S   A++S +           GL++ + + +   T+   + ++WN  
               Sbjct: 70   MVDRLGVDLNVIDVGWPATSSIAWASVVAFFIPLGLIVNVIMLVTKTKT-MNVDIWNF 128

20           Query: 61   FGYMIWGTMAAATGNFILSFAFMVFVLLYSLVMSEVLADRWSEYYGVKNATINSIHNE 120  
                   + Y       + Y   + + I +   V   + +L +++ A   SE+Y +   +I +   I  
               Sbjct: 129   WHYTFMAAVVYTVSDSIIQALIAAVMFQIVALKVADWTAPMVSEFYELPGVSIATGSTIS 188

25           Query: 121   TLIPALILDPLWNLLGVNKVKLNPESLKTKLGIFGEPMTLGFILGVIIGVLGSLRNLASI 180  
                   ++   +   + G+       +P++++ +   GIFGE + +G ILG   IG+L  
               Sbjct: 189   YAPGIWLVKGIQKIPGIKHWNADPDITQRRFGIFGESIFIGLILGAAIGLLAGYNV---- 244

30           Query: 181   DTWGGILGFAVALAAMVTIFPLITGVFASAFAPLAEAVERNKKKESQAEQGALDKKRWFI 240  
                   G ++   +A+AAVM + P + +           P++E+       K           + I  
               Sbjct: 245   ---GEVIEIGMAMAAMVLMPRMVKILMEGLMPVSESAREWLNKR-----FGDREIHI 294

35           Query: 241   AVDDGVGVGEPATIIAGLILVPIMVVISLILPGNEALPVVDLIAIPFMIEAMIAVSKGNI 300  
                   +D V G P+ I   LILVP+ V++++ILPGN   LP DL   IPF++ ++ ++GNI  
               Sbjct: 295   GLDAAVLLGHPSVISTALILVPLTVLLAVILPGNALLPFGDLATIPFIVAFIVGAARGNI 354

              Query: 301   LKAILNGIIFWFSGLYAASALGPIYTEAVKHFGTALPAGVTLIMSFN 347  
                   + ++L G I   +L LY A+ + P++T+ ++       +P G   LI S +  
               Sbjct: 355   IHSVLAGAIMIALSLYMATDIAPVFTKMAENSNFNMPEGSALISSID 401

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1500

40   A DNA sequence (GBSx1587) was identified in *S.galactiae* <SEQ ID 4615> which encodes the amino acid sequence <SEQ ID 4616>. Analysis of this protein sequence reveals the following:

Possible site: 52  
 >>> Seems to have no N-terminal signal sequence

45           ----- Final Results -----  
                   bacterial cytoplasm --- Certainty=0.1013(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50   The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1657-

**Example 1501**

A DNA sequence (GBSx1588) was identified in *S.agalactiae* <SEQ ID 4617> which encodes the amino acid sequence <SEQ ID 4618>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1294(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10147> which encodes amino acid sequence <SEQ ID 10148> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC76604 GB:AE000435 L-xylulose kinase, cryptic [Escherichia coli K12]

Identities = 156/496 (31%), Positives = 261/496 (52%), Gaps = 18/496 (3%)

Query: 16 YYLSIDYGGTNTKALIFDKLGHQIAVSSFETLKNETQSGHRQVNLVKTWNAITSAIREVI 75  
Y+L +D GG+ KA ++D+ G + V Q G + ++ + W + IR ++  
Sbjct: 4 YWLGLDCCGSLKAGLYDREGREAGVQRLPLCALSPQPGWAERDMAELWQCCMAVIRALL 63

Query: 76 QISKLSPEQISAVACIGHGKGLYLLDNKLEPLEQGILSTDNRAKDLAQYFESK--LDNIW 133  
S +S EQI + GKGL+LLD +PL ILS+D RA ++ + ++ + ++  
Sbjct: 64 THSGVSGEQIVGIGISAQKGFLFLDKNDKPLGNAILSSDRRAMEIVRRWQEDGIPEKLY 123

Query: 134 ELTRQHIFPSQSPVILRWLKDYPETYKSIGAVLSAKDFIRYKLTGKVVQEQEYGDASGNHW 193  
LTRQ ++ +LRWLK+++PE Y IG V+ D++R+ LTG E + S ++  
Sbjct: 124 PLTRQTLWTGHPVSLRLRWLKEHEPERYAQIGCVMMTHDYLRWCLTGVKGCEESNISESNL 183

Query: 194 INFQGTGYDPAILDFFGIREIENSLPELIDSADLVPGGISSQAAKETGLVEGTFVVGGLF 253  
N G YDP + D+ GI EI ++LP ++ SA++ G I++Q A TGL GTFVVGGLF  
Sbjct: 184 YNMSLGEYDPCLTDLWLGIAEINHALPPVVGSAEIC-GEITAQTAALTGLKAGTFVVGGLF 242

Query: 254 DIDACALGSGVLESDTFSVISGTWNINT--YPSLKPAKQDSGLMTSYFPDRRYLLEASSP 311  
D+ + AL +G+ + T + + GTW + + L+ + + Y .D. +++ +SP  
Sbjct: 243 DVVSTALCAGIEDEFTLNAVMTWAVTSGITRGLRDGEAHPYVYGRYVNDGEFIVHEASP 302

Query: 312 TSAGNLFNFMKMLMHQEIDNAKSSGGSIYDNLEEFLLTHTDATHHGLIFFPFLYGSNTSQD 371  
TS+GNL + G +D + + + L F PFLYGSN +  
Sbjct: 303 TSSGNLEWF-----TAQWGEISFDEINQAVASLPKAGGDLFFLPFLYGSNAGLE 351

Query: 372 ASACFFGLTTKSTKSQIMIRAVYEGIAFAHKQHITDLIKSRGSPVKIIRFSGGATNSPAWM 431  
++ F+G+ T++ +++A+YEG+ F+H H+ + ++ R + +R +GG +S WM  
Sbjct: 352 MTSGFYGMQAIHTRAHLQAIEYGVVFSHMTHL-NRMRRERFTDVHTLRVTGGPAHSDVWM 410

Query: 432 QMFSDILNFPPIETVEGTELGGGLGAILARHALDKI-SLKEAVQDMVRVKAIYKPQLSEVK 490  
QM +D+ IE + E G G A+ AR + EA +D+ P ++ +  
Sbjct: 411 QMLADVSGLRIELPQVEETGCFGAALAAARVGTGVYHNFSEAQRDLRHPVRTLLPDMTAHQ 470

Query: 491 GYKKKYHAYQKLLLETL 506

Y+KKY YQ L+ L

Sbjct: 471 LYQKKYQRYQHLLAAL 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1658-

**Example 1502**

A DNA sequence (GBSx1589) was identified in *S.agalactiae* <SEQ ID 4619> which encodes the amino acid sequence <SEQ ID 4620>. Analysis of this protein sequence reveals the following:

```

Possible site: 59
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG05648 GB:AE004652 hypothetical protein [Pseudomonas aeruginosa]
Identities = 59/235 (25%), Positives = 104/235 (44%), Gaps = 9/235 (3%)

15 Query: 23 QVQLIKLVKDLGFSRFEIRQELLQDPDRELPAKAEADFYDINLYYSANEDLIK-GGKVN 81
      Q  + L+  G  R  E+R+EL  P  +  AL  A      +  +S+  +L  +  G++N
Sbjct: 23 QASFLPLLAMAGAQRVELREELFAGPP-DTEALTAAIQGLECVFSSPLELWREDGQLN 81

20 Query: 82 PYLNKGLKEASQLGAPFIKLVNGQTRNLSKEELEPLKEILKSQTIGIKVENNQDPKAA TV 141
      P  L  L+  A  GA  ++K+++G  +  +L  L  L  +  +  VEN+Q  P+  +
Sbjct: 82 PELEPTLRRAEACGAGWLKVSGLLPE--QPDLAALGRRLARHGLQLLVENDQTPOGGRI 139

25 Query: 142 ENCQYFMTLVKELQIPISFVFDTANWAFINQDLYQAVNNLACDTTYLHCKNFIQVAGKPH 201
      E  +  F  L  +  Q+  ++  FD  NW  +  Q  +A  L  Y+HCK  I+
Sbjct: 140 EVLERFFRLAERQQDLAMTFDIGNNRWQEQAADEAALRLGRYVG YVHCKAVIRNRD GKL 199

Query: 202 LSKSLFEGEINLTD-LLKSFSNCEYLAL EYPTE----LEILKRDVQRLISISNSQ 251
      ++      ++      LL+  F      A+EYP  +      L  +  +R  +  L  +  Q
30 Sbjct: 200 VAVPPSAADLQYWQRLQLHFPEGVARAIEYPLQGDLLSLRRRIAALARLGQPQ 254

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1503**

A DNA sequence (GBSx1590) was identified in *S.agalactiae* <SEQ ID 4621> which encodes the amino acid sequence <SEQ ID 4622>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0430(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB03939 GB:AP001507 unknown conserved protein [Bacillus halodurans]
Identities = 136/511 (26%), Positives = 234/511 (45%), Gaps = 29/511 (5%)

50 Query: 4 LDKKSYDLLFYLLKLEETVMAIANALNQSRRKVYYHLEKINDALPSDVPQIVSYPRV- 62
      LD++S  +L  LL      +  +  LN  SRR  VY  LEKIN  L      +  V  R
Sbjct: 3 LDQRSTFILTQLLHARSYLP IQELTQKLVNSRRTVYNDLEKINSWLEEQGLKAVYKVR SQ 62

Query: 63 GILLTEKQKAACRLLLDDEVTDYSYVMKSSERLQLSLVSIVVAKDRVTTIDRLMQLNDVSRN 122
      G++L  E+  K      L  +  +  Y  +  ER  ++  ++  +  +  ++  LM  VSRN
55 Sbjct: 63 GLILDERAKEEIPTKLRSLKSWHYEYSAQERKAWVVIYLLTRLEPLFLEHLMDRGTGVSRN 122

```

-1659-

Query: 123 TILNDLNLRLSELAKEYNLQLQSTKCRGYFLDGHPL----SIIQYLYKLLDDIYHNGSS 178  
 T ++D+ L+ EL ++L L+ + GY + G +++ YL + L  
 Sbjct: 123 TTIDDIKCLKDEL--NNFHLALAEFERKDGYYTISGDETDKRKALVYYLSQALPQONWETEL 180

5 Query: 179 SFIDLNFHKLSQLAFGASTYFSKEVLDYFHHYLFISQSLGKKINSQDQGMQILPFI 238  
 S I +F L F+ E L + + S++ L KI D L F+L  
 Sbjct: 181 SPIRIF---LRFKRDNGRIFTIEELQKVYDVISESEKVL--KIQYTDDVLHSLSLRFLLF 235

10 Query: 239 AYRK-----MRLSPEVQTSLSNDFSLVWQRKEYEIAKELADELEENFQLSLDEIEVGLVA 293  
 R +++ P + L KEYE AK ++ +LE+ F + + EV +  
 Sbjct: 236 MKRVAKGKFIKVHPLEKQVLKGT-----KEYEAAKVMSEKLEQAFGVHYPDDEEVLYLT 288

15 Query: 294 MLMLSFRKDRDN-HLESQ-DYDDMRATLTSFLKELEERYHLHFVHKKDLLRQLLTHCKAL 351  
 +LS + + N +ES+ + ++ +TS + + ++ + F K+ L + L H K  
 Sbjct: 289 THILSSKINYANGEIESRKESQELTHIVTSMVNDFOKYACVVFEKELLEKNLFFHIKPA 348

20 Query: 352 LYRKRYGIFSVNPLTEHIKDKYEELFAITSSSVKLEKAWQIKLTDDDVAYLTIHLGGEL 411  
 YR +YG+ N + E IK Y ELF +T V LE+ + D++VA++T+H G +  
 Sbjct: 349 FYRIKYGLEVENNIAESI KTSYPELFLLRKVVHYLERYVGKSVNDNEVAFITMHFVGWM 408

25 Query: 412 RNSQQSPNK-LKLIVIVSDEGIAIQKLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVDM 470  
 R P K K +IV G+ + L Q + DI + +Y+ + VD  
 Sbjct: 409 RREGTIPTRKKALIVCANGVGTSQLKNQLEGLFPAVDIIKTCSIREYEKTP--VEVDF 466

Query: 471 VVSTSDALESRFMLVVPVLTDDDIIRLIR 501  
 ++ST+ E P+ +V+P+LT+ + RL++  
 Sbjct: 467 IISTTSIPEKNVPIFIVNPILTETEKERLLK 497

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4623> which encodes the amino acid  
 30 sequence <SEQ ID 4624>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 35 bacterial cytoplasm --- Certainty=0.0745(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 368/548 (67%), Positives = 456/548 (83%)

Query: 1 MIILDKKSYDLLFYLLKLEPETVMAIANALNQSRKVVYHLEKINDALPSDVPQIVSY 60  
 M+ILDKKSYDLL YLLKLE PETVMAI++ALNQSRKVVY L+KIN ALP V QI+SYP  
 Sbjct: 1 MMILDKKSYDLLSYLLKLETPETVMAISHALNQSRKVVYQLDKINQALPKGVDQIISYP 60

45 Query: 61 RVGILLTEKQKAACRLLLEDEVTDYSYVMKSSERLQSLSVSIVAKDRVTIDRLMQINDVS 120  
 R+GILLT QKAACRLL+EVTDY+YVMKS ER +LS + I V+ +RVTID+LMQ+NDVS  
 Sbjct: 61 RLGILLTADQKAACRLLLEEVTDYNYVMKSDERRRLSSIYIAVSTERVTIDKLMQINDVS 120

50 Query: 121 RNTILNDLNLRLSELAKEYNLQLQSTKCRGYFLDGHPLSIIQYLYKLLDDIYHNGSSSF 180  
 RNTILNDL ELR EL +K+Y +QL +TK RGY+ HP+++IQYLYKLL D+Y G++SF  
 Sbjct: 121 RNTILNDLTELREELEDKQYKIQLHATKARGYYFGCHPMALIQYLYKLLVDVYQGGNTSF 180

55 Query: 181 IDLFNHKLSQLAFGASTYFSKEVLDYFHHYLFISQSLGKKINSQDQGMQILPFI 240  
 ID+FN KLS+ G S YFSK++L YFH YLF+SQ SLGK IN+QD QFM+QILPF+L++Y  
 Sbjct: 181 IDIFNRKLSIQGLSVYFSKDIITYFHEYLFSLQASLGKTINTQDSQFMLQILPFMLLSY 240

60 Query: 241 RKMRLSPEVQTSLSNDFSLVWQRKEYEIAKELADELEENFQLSLDEIEVGLVAMLMLSFR 300  
 R MRL E +++L +F L+W+RKEY IA++LA EL NF+L LD+IEV +VAMLMLSFR  
 Sbjct: 241 RNMRLDSETKSALKQEFHLIWKKEYHIAQDLARELYHNFKLHLLDDIEVSMVAMLMLSFR 300

65 Query: 301 KDRDNHLESQDYDDMRATLTSFLKELEERYHLHFVHKKDLLRQLLTHCKALLYRKRYGIF 360  
 KD+D+H+ESQDYDDMRAT++ F+ +LE RY LHF HK+DLL++L THCKAL+YRK YGIF  
 Sbjct: 301 KDQDHVESQDYDDMRATISHFDIQLESRYQLHFTHKQDLLKRLTTHCKALVYRKAYGIF 360

-1660-

Query: 361 SVNPLTEHIKDKYEELFAITSSSVKLEKAWQIKLTDDDVAYLTIHLGGELRNSQQSPNK 420  
 VNPLT+H+K+KYEELFA+T S +LE+ W I LTDDD+AYLTIHLGGELR++ K  
 Sbjct: 361 LVNPLTDHVKEKYEELFAMTQSCATILEQDWTISLTDDDIAYLTIHLGGELRHNTEQEK 420

5 Query: 421 LKLVIVSDEGIAIQKLLKQCQRYLTNSDIEAVFTTEQYQSVSDLMHVDMMVSTSDALES 480  
 KLVIVSD+GI IQKLL KQCQRYL N IEAVFTTEQYQSV DL+ VDM+V+T+D L++  
 Sbjct: 421 TKLVIVSDDGIGIQKLLFKQCQRYLANGQIEAVFTTEQYQSVYDLLAVDMIVATTDLTKT 480

10 Query: 481 RFPMLVHVHPVLTDDDIIRLIRFSKKGNCANSNQFTNELEKTIAQYVKEDSERYVLKSKIE 540  
 + PML+V+P+L+DDDI+LIRFSK+G + ++F+ EL K I VK++S+RY L SKIE  
 Sbjct: 481 KIPMLIVNPILSDDDIIRLIRFSKQGRLEHSESRFSTELTKAIEAVVKDESRYALVSKIE 540

Query: 541 KLIHQELL 548  
 KLIH+ELL  
 15 Sbjct: 541 KLIHRELL 548

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1504

20 A DNA sequence (GBSx1591) was identified in *S. agalactiae* <SEQ ID 4625> which encodes the amino acid sequence <SEQ ID 4626>. Analysis of this protein sequence reveals the following:

Possible site: 50  
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2692(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC777149 GB:AE000491 orf, hypothetical protein [Escherichia coli K12]  
 Identities = 211/363 (58%), Positives = 270/363 (74%), Gaps = 9/363 (2%)

35 Query: 1 MPNVKIDITRESWILSTFPPEWGTWLNNEEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60  
 M VK ITRESWILSTFPPEWG+WLNEEIE+E VA G FAMWWLG G+W+K+ GG NV +  
 Sbjct: 3 MSKVKISITRESWILSTFPPEWGSWLNEEIEQEQVAPGTFAMWWLGCTGIWLKSEGGTNVCV 62

40 Query: 61 DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYVLVSHFHS 120  
 D W GK + M +GHQM MAGV+KLQPNLR P V+DPFAI ++D L +H H+  
 Sbjct: 63 DFWCCTGKQSHGNPLMKQGHQMORMAGVKKLQPNLRRTTFFVLDPFAIRQIDAVLATHDHN 122

45 Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGEIWKWGVPEERIIVIKPGESFEFKDIKVT 180  
 DHID+N AAA++ N D V F+GP C ++W WGVP+ER IV+KPG+ + KDI++ A  
 Sbjct: 123 DHIDVNVAAMQNC-ADDVFFIGPKTCVDLWIGWGVKERCIVVKPGDVVKVDIEIHA 181

Query: 181 VESFDRTCLVTLPLVDGAEHHDGELAGLAVTDEEMARKAVNYIFETPGGTIYHGADSHFSN 240  
 +++FDRT L+TLP D + AG V + M +AVNY+F+TPGG++YH DSH+SN  
 Sbjct: 182 LDAFDRTALITLPAQ-----KAAG--VLPDGMDDRAVNYLFKTPGGSILYHSGDSHYSN 233

50 Query: 241 YFAKHGKDYKIDVAINNYGDNVPVGIQDKMTSIDLLRMAENLRKVIIIPVHYDIWSNFMAS 300  
 Y+AKHG +++IDVA+ +YG+NP GI DKMTS D+LRM E L AKV+IP H+DIWSNF A  
 Sbjct: 234 YYAKHGNEHQIDVALGSYGENPRGITDKMTSADMLRMGEALNAKVVIIPFHDIWSNFDQAD 293

55 Query: 301 TDEILQLWKMRKERLQYDFHFFIWEVGGKYTPQDKDRIEYHHPRGFDDCFEQESNIQFK 360  
 EI LW+M+K+RL+Y F PFIW+VGGK+T+P DKD EYH+PRGFDDCF E ++ FK  
 Sbjct: 294 PQEIRVLWEMKKDRKLYGFKPFIWQVGGKFTWPLDKDNFEYHYPRGFDDCFETIEPDLFPK 353

Query: 361 ALL 363  
 + L  
 60 Sbjct: 354 SFL 356

-1661-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4627> which encodes the amino acid sequence <SEQ ID 4628>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3298(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 315/363 (86%), Positives = 348/363 (95%)

Query: 1 MPNVKDITRESWILSTFPWGTWLNNEEIEEEVVAEGNFAMWWLGNCGVWIKTPGGANVVM 60  
M V+DITRESWIL+TFPEWGTWLNNEEIE+EVV NFAMWWLGNCG+WIKTPGGANVVM  
Sbjct: 1 MTKVQDITRESWILNTFPWGTWLNNEEIEQEVVPADNFAMWWLGNCGIWIKTPGGANVVM 60

Query: 61 DLWSNRGKSTKKVKDMVRGHQMANMAGVRKLQPNLRAQPMVIDPFAINELDYLLVSHFHS 120  
DLWSNRGK+TK+VKDMVRGHQMANMAG RKLQPNLRAQPMVIDPF INELDYLLVSH+HS  
Sbjct: 61 DLWSNRGKATKQVKDMVRGHQMANMAGARKLQPNLRAQPMVIDPFMINELDYLLVSHYHS 120

Query: 121 DHIDINTAAAIINNPNLDHVKFVGPYECGEIWKKGVPPEERIIIVIKPGESFEFKDIKITA 180  
DHIDINTAAAIINNP L+HVKFVGPYECGE+WK WGV++RI+++KPG+SFEFKDIK+TA  
Sbjct: 121 DHIDINTAAAIINNPKLNHVKFVGPYECGEVWKNWGVPKDRIMILKPGDSFEFKDIKITA 180

Query: 181 VESFDRTCCLVTLPLVDGAEHGDGELAGLAVTDEEMARKAVNYIFETPGGTIYHGADSHFSN 240  
VESFDRTCCLVTLPL+ GA+ DG+LAGLA+TD++MARKAVNYIFETPGGTIYHGADSHFSN  
Sbjct: 181 VESFDRTCCLVTLPLIQGADAQDGLAGLAITDDDMARKAVNYIFETPGGTIYHGADSHFSN 240

Query: 241 YFAKHGKDYKIDVAINNYGDNPNVGIQDKMTSIDLLRMAENLRAKVIIIPVHYDIWSNFMAS 300  
YFAKHG+DY IDV +NNYG+NP+GIQDKMTS+DLLRMAENLRAKV+IPVHYDIWSNFMAS  
Sbjct: 241 YFAKHGRDYDIDVVLNNYGENPIGIQDKMTSVDLLRMAENLRAKVVIIPVHYDIWSNFMAS 300

Query: 301 TDEILQLWKMRLQYDFHPPFIWEVGGKYTPQDKDRIEYHHPRGFDDCFEQESNIQFK 360  
TDEIL+LWKMRLQYDFHPPFIWEVGGKYTPQD++RIEYHHPRGFDDCF ++SNIQFK  
Sbjct: 301 TDEILELWKMRLQYDFHPPFIWEVGGKYTPQDQNRIEYHHPRGFDDCFLEDSNIQFK 360

Query: 361 ALL 363

ALL

Sbjct: 361 ALL 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1505

A DNA sequence (GBSx1592) was identified in *S.agalactiae* <SEQ ID 4629> which encodes the amino acid sequence <SEQ ID 4630>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3988(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10145> which encodes amino acid sequence <SEQ ID 10146> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1662-

&gt;GP:BAA18808 GB:D90917 hypothetical protein [Synechocystis sp.]

Identities = 358/785 (45%), Positives = 494/785 (62%), Gaps = 15/785 (1%)

5 Query: 22 LEKLDANWRAANYISAAQMYLKDNPLRLRELVDLKVHPIGHWGTVPQGNFIYAHNLRA 81  
L ++ +WRAANY++ +YL+DNPLLR L +K +GHWG+ PG +F+Y HLNLR  
Sbjct: 44 LNQMHGFWRAANYLAVGMIYLRDNPLRLREPLQPEQIKHRLLGHWGSSPGISFLYTHLNRI 103

10 Query: 82 INKYDLDMFYIEGPGHGGQVMVSN SYLDG SYTELNP NIEQTE DGF KQLCKIFSFPCGIAS 141  
I K+D DM Y+ GPGHG + YL+GSY+ + EDG K+ K FSFP GI S  
Sbjct: 104 IRKFDQDMLYMGPGHGAPGFLGPCYLEGSYSRFFAECSEDEDCMKRFFKQFSFPGGIGS 163

15 Query: 142 HAAPETPGSIHEGGELGYALSHATGAILDNP DVIAATVIGDGEGETGPLMAGWLSNTFIN 201  
H PETPGSIHEGGELGY LSHA GA DNP++I + GDGE ETGPEL W SN FIN  
Sbjct: 164 HCTPETPGSIHEGGELGYCLSHAYGAAFDNP NLIVVGLAGDGESETGPLATSWHSNKFIN 223

20 Query: 202 PVNDGAVLP I FYLNGGKIHNPTIFERKTDEELSQFF EGLGWKPIFADVV ELS EDAHAAA 261  
P+ DGAVLP+ +LNG KI+NP++ R + EEL FEG G+ P F + D + H  
Sbjct: 224 PIRDGAVLPVLHLNGYKINNPSVLSRISHEELKALFEGYGYTPYFVE---GSDPESMHQ 279

25 Query: 262 LFAEKLDQAIQEIKTIQSEARQKPAEEAIQAKFPVLVARIPKGWTGPKAWEGTPIEGGFR 321  
A LD + EI IQ EAR A++ ++P++V R PKGWTGP +G +EG +R  
Sbjct: 280 AMAATLDHCVSEIHQIQEARESTGI--AVRPRWPMVMVRTPKGWTGPDYVDGHKVEGFWR 337

30 Query: 322 AHQVPIPVDAHHMEHVDLSLSWLQSYRPEELFDENGKIVDEIAAISPKGDRRMSMNPITN 381  
+HQVP+ + H+ L +W++SY+PEELFDE G + AI+P+GD+R+ P N  
Sbjct: 338 SHQVPMGGMHENPAHLQQLEAWMRSYKPEELFDEQGTLPKPGFKALAEPEGDKRLGSTPYAN 397

35 Query: 382 AGIV-KAMDTADWKKFALDINVPQIQMAQDMIEFGKYAADLVDANPDNFRIFGPDET KSN 440  
G++ + + D++++ +D++ PG I A + G + D++ N NFR+FGPDE SN  
Sbjct: 398 GGLLRRLKMPDFRQYQIDVDQPGTIEAPNTAPLGVL RDVMANNMTNFR LFGPDENSSN 457

40 Query: 441 RLQEVFTRTSRQWLGRRKPDYDEA--LSPAGRVIDSQLSEHQAEGLFEGYVLTGRHGFFA 498  
+L V+ ++ W+ + + LSP GRV++ LSEH EG+LE Y+LTGRHGFFA  
Sbjct: 458 KLHAVVEVSKKFWIAEYLEEDQDGELSPDGRVME-MLSEHTLEGWLEAYLLTGRHGFFA 516

45 Query: 499 SYESFLRVVDSMVTOHFQKWLRSKTHTTWRKNYPALNLIAASTVFQQDHNGYTHQDPGIL 558  
+YESF V+ SMV QH KWL + H WR + +LN++ STV++QDHNG+THQDPG L  
Sbjct: 517 TYESFAHVITSMVNQHAKWLDICR-HLNWRADISSLNILMTSTVWRQDHNGFTHQDPGFL 575

50 Query: 559 THLAETPEYIREYLPADTNSLLAVMDKAFKAEDKINLIVTSKHPRPQFYSAEAEELVA 618  
+ K+P+ +R YLP D NSLL+V D ++++ IN+IV K Q+ + A  
Sbjct: 576 DVILNKSPDVVRIYLPDPVNSLSVADHCLQSKNYINIIVCDKQAHLYQYQDMTSAIRNCT 635

55 Query: 619 EGYKVIDWASNVSLNQEPDVVFAAAGTEPNLEALAAISILHKAPPELKIRFVNVDILKL 678  
+G + +WASN EPD VV AAAG P EALAA ++L + FP L+IRFV+V+D+LKL  
Sbjct: 636 KGVDIWEWASN-DAGTEPDVVMMAAGDIPTKEALAAATAMLRQFFPNLRIRFVSVIDLLKL 694

60 Query: 679 RHPSQDARGLSDEEFNKVFTTDKPVIFAFHGYEDMIRDIFFSRHNH-NLHTHGYRENGDI 737  
+ S+ GLSD +F+ +FTTDKP+IF FH Y +I + + R NH NLH GY+E G+I  
Sbjct: 695 QPESEHPHGLSDRDFDSLFTTDKPIIFNFHAYPWLHRLTYRRTNHNHNLHVRGYKEKGN 754

65 Query: 738 TTPFDMRVMSSELDRFHQAQDA--ALASLGNKAQAFSDENMQMVAYHKDYIREHGDDIPEV 795  
TP D+ + +++DRF LA D L L + + M +Y EHG D+PE+  
Sbjct: 755 NTPMDLAIQNQIDRFLAIDVIDRLPQLRVAGAHKEMLKDMQIDCTNYAYEHGIDMPEI 814

70 Query: 796 QNWKW 800  
NW+W  
Sbjct: 815 VNNRW 819

60 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1663-

**Example 1506**

A DNA sequence (GBSx1593) was identified in *S.agalactiae* <SEQ ID 4631> which encodes the amino acid sequence <SEQ ID 4632>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 32
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3509(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF37878 GB:AF234619 OpuAA [Lactococcus lactis]
Identities = 274/402 (68%), Positives = 338/402 (83%)

15  Query: 5  LEVKNLTKIFGKKQKAALFEMVKQKSKTEILEKTGATVGVDASFEIKEGEIFVIMGLSG 64
      +++++LTKIFGK+ K AL MV++G+ K EIL+KTGATVGVDYD +FEI EGEIFVIMGLSG
      Sbjct: 5  IKIEHLTKIFGKRIKTALTMVEKGEPKNEILKKTGATVGVDYDINFEINEGEIFVIMGLSG 64

20  Query: 65  SGKSTLVRMLNRLIDPSSGNIYLDGKDIAMNVEDLRNIRRH DINMVFNFGFLFPHRTIL 124
      SGKSTL+R+LNRLI+P+SG I++D +D+A +N EDL +RR ++MVFNFGFLFPHRTIL
      Sbjct: 65  SGKSTLLRLNRLIEPTSGKIFIDNQDVATLNKEDLLQVRRKTM SMVFQNFGLFPHRTIL 124

25  Query: 125  ENTEFGLFMRGVSKERTTLAEKALDNAGLLPFKDQYPSQLSGGMQQRVGLARALANSPK 184
      ENTE+GLE++ V KEER AEKALDNA LL FKDQYP QLSGGMQQRVGLARALAN P+
      Sbjct: 125  ENTEYGLEVNVPKEERRKRAEKALDNANLLDFKDQYPKQLSGGMQQRVGLARALANDPE 184

30  Query: 185  ILLMDEAFSALDPLIRREMQDELLDLQDTNKTIIIFSHDLNEALRIGDRIALMKDGEIM 244
      ILLMDEAFSALDPLIRREMQDELL+LQ ++TIIF+SHDLNEALRIGDRIA+MKDG+IM
      Sbjct: 185  ILLMDEAFSALDPLIRREMQDELLLELQAKFQKTIIFVSHDLNEALRIGDRIAIMKDGKIM 244

35  Query: 245  QIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTVLEIDGQVALTRMHREE 304
      QIGTGEEILTNPAND+V+ FVEDVDR+KV+TA+NIMI LTT +++DGP VAL +M EE
      Sbjct: 245  QIGTGEEILTNPANDYVKTFFVEDVDRKAVITAENIMIPALTTNIDVDGPSVALKKMKTEE 304

40  Query: 305  VSMLMATNRRRQLLGSALTADAAIEARKKDLPLSEVIDKDVVTVSKDTVITDIMPLIYDSS 364
      VS LMA +++RQ G +T++ AI ARK + PL +V+ DV TVSK+ ++ DI+P+IYD+
      Sbjct: 305  VSSLMVAVDKKRQFRGVVTSQAIAARKNNQPLKDVMTTDVGTVSKEMLVRLDILPIIYDAP 364

40  Query: 365  APIAVTDDNDRLLGVIIIRGRVIEALANVQDET VVESPKEITVE 406
      P+AV DDN L GV+IRG V+EALA++ DE VE ++ E
      Sbjct: 365  TPLAVVDDNGFLKGVLRGSLVLEALADIPDEDEVEEIEKEEE 406

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4633> which encodes the amino acid sequence <SEQ ID 4634>. Analysis of this protein sequence reveals the following:

```

      Possible site: 53
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
50     bacterial cytoplasm --- Certainty=0.3761(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

55  Identities = 344/395 (87%), Positives = 374/395 (94%)

      Query: 1  MTNILEVKNLTKIFGKKQKAALFEMVKQKSKTEILEKTGATVGVDASFEIKEGEIFVIM 60
      M ILEVK+L+KIFGKKQKAALFEMVK GK+K+EI +KTGATVGVDASFE+K+GEIFVIM
      Sbjct: 1  METILEVKHLSKIFGKKQKAALFEMVKTKGNKSEIFKKTGATVGVDASFEVKKGEIFVIM 60

60  Query: 61  GLSGSGKSTLVRMLNRLIDPSSGNIYLDGKDIAMNVEDLRNIRRH DINMVFNFGFLFPH 120

```

-1664-

GLSGSGKSTLVRMLNRLI+PS+G+I L+GKDI+ M+ + LR +RRHDINMVFO+F LFPH  
 Sbjct: 61 GLSGSGKSTLVRMLNRLIEPSAGSILLEGKDISTMSADQLREVRRHDINMVFSFALFPH 120  
 Query: 121 RTILENTEFGLEMRGVSKERTTTAEKALDNAGLLPFKDQYPSQLSGGMQQRVGLARALA 180  
 +TILENTEFGLE+RGV KEER LAEKALDN+GLL FKDQYP+QLSGGMQQRVGLARALA  
 Sbjct: 121 KTIENTEFGLELRGVPEERQRLAEKALDNSGLLDFKDQYPNQLSGGMQQRVGLARALA 180  
 Query: 181 NSPKILLMDEAFSALDPLIRREMDELDDLDQD'TNKQTIIFISHDLNEALRIGDRIALMKD 240  
 NSPKILLMDEAFSALDPLIRREMDELDDLDQD+ KQTIIFISHDLNEALRIGDRIALMKD  
 Sbjct: 181 NSPKILLMDEAFSALDPLIRREMDELDDLDQDSMKQTIIFISHDLNEALRIGDRIALMKD 240  
 Query: 241 GEIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTVLEIDGPQVALTRM 300  
 G+IMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTT +E+DGPQVAL RM  
 Sbjct: 241 GQIMQIGTGEEILTNPANDFVREFVEDVDRSKVLTAQNIMIKPLTTTVELDGPQVALNRM 300  
 Query: 301 HREEVSMATNRRRQLLGSALTADAAIEARKKDLPLSEVIDKDVTVSKDTVITDILPLI 360  
 H EEVSMATNRRRQL+GSLTADAAIEARKK LPLSEVID+DV TVSKDT+ITDI+PLI  
 Sbjct: 301 HNEEVSMATNRRRQLVGSALTADAAIEARKKGLPLSEVIDRDVRTVSKDTIITDILPLI 360  
 Query: 361 YDSSAPIAVTDDNDRLLGVIIRGRVIEALANVQDE 395  
 YDSSAPIAVTDDN+RLLGVIIRGRVIEALAN+ DE  
 Sbjct: 361 YDSSAPIAVTDDNDRLLGVIIRGRVIEALANISDE 395

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1507

A DNA sequence (GBSx1594) was identified in *S.agalactiae* <SEQ ID 4635> which encodes the amino acid sequence <SEQ ID 4636>. This protein is predicted to be OpuABC (opuAB). Analysis of this protein sequence reveals the following:

Possible site: 41  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -10.67 Transmembrane 48 - 64 ( 43 - 72)  
 INTEGRAL Likelihood = -9.24 Transmembrane 101 - 117 ( 93 - 122)  
 INTEGRAL Likelihood = -7.54 Transmembrane 296 - 312 ( 290 - 316)  
 INTEGRAL Likelihood = -6.21 Transmembrane 252 - 268 ( 250 - 273)  
 INTEGRAL Likelihood = -5.57 Transmembrane 141 - 157 ( 138 - 170)  
 INTEGRAL Likelihood = -0.53 Transmembrane 220 - 236 ( 220 - 237)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5267(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GF:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]  
 Identities = 345/578 (59%), Positives = 429/578 (73%), Gaps = 8/578 (1%)  
 Query: 1 MENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQITGNALMDWMTKTLFINPILFIVLI 60  
 M+L ++P+A +V S T+WIT TFS FD IQ G LM+ +T L + L I ++  
 Sbjct: 1 MIDLAIGQVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWLMIAVV 60  
 Query: 61 TIAVFFLAKKKQLPTFTFFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120  
 TI ++ KK P FTFIGL I NQGLW L++T LVL++SL+SIIIGVPLGIWMA  
 Sbjct: 61 TILAILVSGKKIAFPFLFTFIGLSLIANQGLWSDLMSTTTLVLLSSLLSIIIGVPLGIWMA 120  
 Query: 121 KSDKVKQVVPILDFMQTMPAFVYLIPAVAFFGIGMVGVFASVVFALPPTVRFTNLAIR 180  
 KSD V ++V PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASV+FALEPPTVR TNL IR  
 Sbjct: 121 KSDLVAKIVQPIILDFMQTMPGFVYLIPAVAFFGIGVVGVFASVIFALPPTVRMTNLGIR 180  
 Query: 181 EIPLELIEASDSFGSTVKQKFLKVELPLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240  
 ++ EL+EA+DSFGST +QKLFK+E PLAK TIMAG+NQT+MLALSMVV SMIGAPGLG

-1665-

Sbjct: 181 QVSTELVEADSFSGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMGAPGLG 240

Query: 241 REVL SALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGKEQAKTSKVKKW---VGLG 297  
R VL+A+Q ADIG GFVSG+SLVILAI++DR +Q N P EKQ + VKKW + L

5 Sbjct: 241 RGVLAAVQSADIGKGFVSGISLVILAIIDRFTQKLNVS PLEKQGNPT-VKKWKRGIALV 299

Query: 298 ALALFILAAALGRIVVNMTSGNEAKGQKVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTP 357  
+L I+ A M+ G A +KV + Y+ WDSEVAS NV+ + +K G+DV+ T

10 Sbjct: 300 SLIALIIGAFS---GMSFGKTASDKKVDLVYMNWDSEVASINVLTQAMKEHGFVDKTTA 355

Query: 358 LDNAVMMQTVANGNADFTTSAWLPKTHGQYFNKYKNSLDDLGPHEVNVKIGLVVPKYMNV 417  
LDNAV WQTVANG AD SAWLP TH + KY S+D LGP+++ K+G VVP YMN

Sbjct: 356 LDNAVAWQTVANGQADGMVSAWLPNTHKTQWQKYGKSVDDLGPNLKGAKGVFVVPYMN 415

15 Query: 418 NSIEELSNOADKQITGIEPGAGIMKSAQSLKDYPNLSSWKL SASTGAMTTTLGKAIGN 477  
NSIE+L+NQA+K ITGIEPGAG+M +++++L Y NL WKL+ +S+GAMT LG+AIK

Sbjct: 416 NSIEDLTNQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSSGAMTVALGEAIKQ 475

Query: 478 KDQVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKDKMPKVYKIIDKFKW 537  
+VITGWSPHWMF KYDLKYL DPK + G E+INTI RK LKK+ P+ YK++DKF W

20 Sbjct: 476 HKDIVITGWSPHWMFNKYDLKYLADPKGTMTSENINTIVRKGLKENPEAYKVLDFKNW 535

Query: 538 TKEDMESIMLMDKGM EPAKAAQKWIKNHKKEVSEWTK 575  
T +DME++MLD+ G P +AA+ WIK+H+KEV +W K

25 Sbjct: 536 TTKDMEAVMLDIQNGKTP EEAKNWIKDHQKEVDKWF 573

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4637> which encodes the amino acid sequence <SEQ ID 4638>. Analysis of this protein sequence reveals the following:

Possible site: 47

30 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.86	Transmembrane	101 - 117 ( 93 - 121)
INTEGRAL	Likelihood = -7.54	Transmembrane	252 - 268 ( 250 - 273)
INTEGRAL	Likelihood = -6.85	Transmembrane	48 - 64 ( 43 - 70)
INTEGRAL	Likelihood = -5.57	Transmembrane	141 - 157 ( 138 - 170)
35 INTEGRAL	Likelihood = -5.26	Transmembrane	295 - 311 ( 289 - 315)
INTEGRAL	Likelihood = -0.53	Transmembrane	220 - 236 ( 220 - 237)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4545(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:AAF37879 GB:AF234619 OpuABC [Lactococcus lactis]  
Identities = 340/571 (59%), Positives = 418/571 (72%), Gaps = 8/571 (1%)

Query: 8 KLPVAQLVEQLTEWLTKTFSGLFDIMQVVSFLMDWMTKTLFIHPLLFIVLVTAGMFFL 67  
++P+A V T+W+T TFS FD++Q G+ LM+ +T L + L I +VT +

50 Sbjct: 8 QVPIANWVSSATDWITSTFSSGFDVIQKSGTVLMNGITGALTAVPFWLMIAVVITILAILV 67

Query: 68 AKKKWPLPTFTLLGLLFIYNQGLWKQLMNTFTLVLVASLISVLIGIPLGIWMAKNATVRQ 127  
+ KK P FT +GL I NQGLW LM+T TLVL++SL+S++IG+PLGIWMAK+ V +

Sbjct: 68 SGKKIAPPLFTFTIGLSLIANQGLWSDLMSTITLVLLSLLSIIGVPLGIWMAKSDLVAK 127

55 Query: 128 IVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIIRDIPTELI 187  
IV PILDFMQTMP FVYLIPAVAFFGIG+VPGVFASVIFALPPTVR TNL IR + TEL+

Sbjct: 128 IVQPILDFMQTMPGFVYLIPAVAFFGIGVVPGVFASVIFALPPTVRMTNLGIRQVSTELV 187

60 Query: 188 EASDAFGSTGKQKLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLGREVLSAL 247  
EA+D+FGST +QKLFK+E PLAK TIMAGVNQT+MLALSMV SMIGAPGLGR VL+A+

Sbjct: 188 EAADSFSGSTARQKLFKLEFPLAKGTIMAGVNQTIMLALSMVVIASMGAPGLGRGVLA 247

Query: 248 QHADIGSGFVSGLALVILAIVLDRMTQLFNSKPGKEKAKAGKTNKW---IGLAALAVFLIA 304  
Q ADIG GFVSG++LVILAI++DR TQ N P EK KW I L +L +I

65 Sbjct: 248 QSADIGKGFVSGISLVILAIIDRFTQKLNVS PLEKQGNPTVKKWKRGIALVSLALIIIG 307

-1666-

5 Query: 305 ALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTLTPLDNAVMWQ 364  
 A M+ G + V++ Y+ WDSEVAS +V+ + +K G+ V T LDNAV WQ  
 Sbjct: 308 AFS---GMSFGKTASDKKVDLVYMNWDSEVASINVLQAMKEHGFDVKTTALDNAVAWQ 363

10 Query: 365 TVANGNADFSTSAWLPVTHGQQYQKYKSKLDDLGPNLKGTGLGLAVPKYMTDVNSIEDLS 424  
 TVANG AD SAWLP TH Q+QKY +D LGPNLKG K+G VP YM +VNSIEDL+  
 Sbjct: 364 TVANGQADGMVSAWLPNTHKTQWQKYGKSVDDLGPNLKGAKVGFVPSYM-NVNSIEDLIT 422

15 Query: 425 KQADQKITGIEPGAGIMAAQAQKTLKEYHNLSSWELVAASTGAMTTSLDQAIAKKKDPPIVVT 484  
 QA++ ITGIEPGAG+MAA++KTL Y NL W+LV +S+GAMT +L +AIK+ IV+T  
 Sbjct: 423 NQANKTITGIEPGAGVMAASEKTLNSYDNLKDWKLVPSSSGAMTVALGEAIKQHKDITVIT 482

20 Query: 485 AWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWTKDMEA 544  
 WSPHWMF KYDLKYL DPK G++ENINTI RKGLKKE P YK++DKF+WT KDMEA  
 Sbjct: 483 GWSPHWMFNKYDLKYLADPKGTMTSENINTIVRKGLKKENPEAYKVLDFNWTTKDMEA 542

Query: 545 VMLDINKGMSPEAAAKWVEANKSKVSSWTK 575  
 VMLDI G +PE AAK W++ ++ +V W K  
 Sbjct: 543 VMLDIQNGKTPEEAAKNWIKDHQKEVDKWF 573

An alignment of the GAS and GBS proteins is shown below.

Identities = 439/576 (76%), Positives = 513/576 (88%), Gaps = 2/576 (0%)

25 Query: 1 MENLLQHKLPVAPFVESTTNWITKTFSGLFDFIQTIGNALMDWMTKTLFFINPLLFIVLI 60  
 +E +LQ KLPVA VE T W+TKTFSGLFD +Q +G+ LMDWMTKTLFFI+PLLFIVL+  
 Sbjct: 1 LETILQTKLPVAQLVEQLTEWLTKTFSGLFDIMQVVSFLMDWMTKTLFFIHPPLLFIVLV 60

30 Query: 61 TIAVFFLAKKKWQLPTFTFTFIGLLFIYNQGLWEQLINTFNLVLVASLISIIIGVPLGIWMA 120  
 T +FFLAKKKW LPTFT +GLLFIYNQGLW+QL+NTF LVLVASLIS++IG+PLGIWMA  
 Sbjct: 61 TAGMFFLAKKKWPLPTFTTLLGLLFIYNQGLWQLMNTFTLVLVASLISVLIGIPLGIWMA 120

35 Query: 121 KSDKVKQVVPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVVFALPPTVRFTNLAIR 180  
 K+ V+Q+VNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASV+FALEPPTVRFTNLAIR  
 Sbjct: 121 KNATVRQIVNPILDFMQTMPAFVYLIPAVAFFGIGMVPGVFASVIFALPPTVRFTNLAIR 180

40 Query: 181 EIPLELIEASDSFGSTVKQLFKVELPLAKNTIMAGINQTMMLALSMVVTGSMIGAPGLG 240  
 +IP ELIEASD+FGST KQKLFKVELPLAKNTIMAG+NQTMMLALSMVVTGSMIGAPGLG  
 Sbjct: 181 DIPTLIEASDAFGSTGKQLFKVELPLAKNTIMAGVNQTMMLALSMVVTGSMIGAPGLG 240

45 Query: 241 REVLALQHADIGTGFVSGLSLVILAIVLDRVSQFFNSKPGEKQAKTSKVKKWVGLGALA 300  
 REVLALQHADIG+GFVSGL+LVILAIVLDR++Q FNSKP EK AK K KW+GL ALA  
 Sbjct: 241 REVLALQHADIGSGFVSGLALVILAIVLDRMTQLFNSKPQEK-AKAGKTNKWIGLALA 299

50 Query: 301 LFILAALGRIVVNMTSGNEAKGQVKIAYVQWDSEVASTNVIAEVLKSKGYDVELTPLDN 360  
 +F++AALGR ++ MTSG K+ V IAYVQWDSEVAST+VIAEVLK++GY V LTPLDN  
 Sbjct: 300 VFLIAALGRGIMAMTSGMADKGETVNIAYVQWDSEVASTHVIAEVLKNEGYHVTLTPLDN 359

55 Query: 361 AVMWQTVANGNADFTTSAWLPKTHGQYFNKYKNSLDDLGPVENVKIGLVVPKYM-NVNS 419  
 AVMWQTVANGNADF+TSAWLP THGQ + KYK+ LDDLGP+++ K+GL VPKYM +VNS  
 Sbjct: 360 AVMWQTVANGNADFSTSAWLPVTHGQQYQKYKSKLDDLGPNLKGTGLGLAVPKYMTDVNS 419

60 Query: 420 IEELSNQADKQITGIEPGAGIMSAKQSLKDYPNLSSWKLASASTGAMTTTLGKAIAKND 479  
 IE+LS QAD++ITGIEPGAGIM +A+++LK+Y NLSSW+L+++ASTGAMTT+L +AIK KD  
 Sbjct: 420 IEDLSKQADQKITGIEPGAGIMAAQAQKTLKEYHNLSSWELVAASTGAMTTSLDQAIAKKD 479

Query: 480 QVVITGWSPHWMFAKYDLKYLKDPKKSFGGEEHINTIARKNLKKDMPKVYKIIDKFHWTK 539  
 +V+T WSPHWMFAKYDLKYLKDPK+ FG E+INTIARK LKK++P VYKIIDKF WT+  
 Sbjct: 480 PIVVTAWSPHWMFAKYDLKYLKDPKEIFGSTENINTIARKGLKKELPNVYKIIDKFHWTK 539

65 Query: 540 EDMESIMLDMDKGMPEAKAAQKWIKNHKKEVSEWTK 575  
 +DME++MLD++KGM P AA+KW++ +K +VS WTK  
 Sbjct: 540 KDMEAVMLDINKGMSPEAAAKWVEANKSKVSSWTK 575

65 A related GBS gene <SEQ ID 8827> and protein <SEQ ID 8828> were also identified. Analysis of this protein sequence reveals the following:



-1668-

```

      |||: :||| ||: ||| :||| ||| ||| ||| ||| : | |: ||| || |||: |: ||: |||
DWKLVPSSSGAMFVALGEAIKQHKDIVITGWSPHWMFNKYDLKYLADPKGTMGTSENINTIVRKGLKKENPEAYKVLDKF
      470      480      490      500      510      520      530

5      1935      1965      1995      2025      2055      2085      2115      2145
      KWTKEDMESIMLDMDKGMEPAKAAQKWKIKNHKKEVSEWTK*YRKKHVSFRACFLM*LKSF*LFNISFILF*YIKSERMKE
      || :|||: |||: | | :||: |||: |||: ||| :| |
      NWTTKDMEAVMLDIQNGKTPEEAAKNWIKDHQKEVDKWFK
      550      560      570

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1508

A DNA sequence (GBSx1596) was identified in *S. agalactiae* <SEQ ID 4639> which encodes the amino acid sequence <SEQ ID 4640>. This protein is predicted to be a transposase. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.65      Transmembrane      223 - 239 ( 223 - 240)

----- Final Results -----
      bacterial membrane --- Certainty=0.1659(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10057> which encodes amino acid sequence <SEQ ID 10058> was also identified. A related GBS nucleic acid sequence <SEQ ID 10031> which encodes amino acid sequence <SEQ ID 10032> was also identified. A related GBS nucleic acid sequence <SEQ ID 10801> which encodes amino acid sequence <SEQ ID 10802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA50689 GB:X71844 putative transposase [Clostridium perfringens]
Identities = 94/364 (25%), Positives = 160/364 (43%), Gaps = 35/364 (9%)

Query: 8   KHKHLTLLDRNDIQSGLDRGETFKAIGLNLKHPTTIAKEVKRN--KQLRESTKDCLDCP 65
          K+KHL + +R ++ L G +      L + T+ E++R KQ+++ + +
Sbjct: 12  KNKHLNMKERMIVEIRLKDGFSAKYNTKELNRPINTVLNEIRRGTTKQIKQKQKEFHVYFA 71

Query: 66  LLRKAPYVCNGCPKRRINCGYKKTFFYLAKQAQRNYEKLIVSREGIPLNKETFWKIDRVL 125
          +A Y N + + N YK      ++ K +V+      K W +D +
Sbjct: 72  DTGEAVYKKN---RLKSNRKYKLL-----ECSDFIKYVVDKV-----KNDHWSLDACV 116

Query: 126 SNGVKKQQRIRYHILKTNDLEVSSSTVYRHIKGYLSITPIDLPRAVKFKRRKSTLPPIP 185
          G+ ++ +      +S+ T+Y ++ G L I IDLP K + +KST
Sbjct: 117 -----GEALHSSRFSPSQIISTKTLNYVDLGLLPIKNIDLP--AKLHRNKKSTVRVNN 168

Query: 186 KAIKEGRRYEDFIEHM-NQSELNSWLEMDTVIGRIGGK--VLLTFNVAFCNFIFAKLMDS 242
          K K G D + N+ E W E+D V+G K VLLT + M S
Sbjct: 169 KK-KLGTSISDRPNSIENREEFGHW-EIDCVLGEKSNKDKVLLTLVERKTRYAIISEMSS 226

Query: 243 KTAIETAKHIQVIKRTLYDNKRDFELFPVILTNGGEFARVDDIEIDVCGQSOLFCDP 302
          + I K + IK L      F E+F I DNG EFA + + E+ +++++F P
Sbjct: 227 HSTISVTKALDKIKEFLGSK---PSEVFKSITADNGSEFADLSEFELKT--KTKVYFTHP 281

Query: 303 NRSQKARIEKNHTLVRDILPKGTSFDNLTQEDINLALSHINSVKRQALNGKTAYELFSF 362
          S +K E+++ L+R +PKG + + E I+ + +N++ R+ L+ KT ELF
Sbjct: 282 YSSFEGKTNERHNLIRRFIPKGRISDYSLETISFIENWMNTLPRKLLDYKTPEELFEI 341

Query: 363 TYGK 366

```

-1669-

K  
Sbjct: 342 HLDK 345

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1509

A DNA sequence (GBSx1597) was identified in *S.agalactiae* <SEQ ID 4641> which encodes the amino acid sequence <SEQ ID 4642>. Analysis of this protein sequence reveals the following:

10 Possible site: 33  
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -11.30	Transmembrane	56 - 72 ( 48 - 79)
INTEGRAL	Likelihood = -6.85	Transmembrane	11 - 27 ( 6 - 30)
INTEGRAL	Likelihood = -6.69	Transmembrane	129 - 145 ( 126 - 158)
15 INTEGRAL	Likelihood = -6.53	Transmembrane	94 - 110 ( 90 - 117)
INTEGRAL	Likelihood = -1.54	Transmembrane	216 - 232 ( 215 - 232)
INTEGRAL	Likelihood = -1.22	Transmembrane	147 - 163 ( 147 - 165)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- A related GBS nucleic acid sequence <SEQ ID 9431> which encodes amino acid sequence <SEQ ID 9432>  
25 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07666 GB:AP001520 unknown conserved protein [Bacillus halodurans]  
Identities = 112/224 (50%), Positives = 150/224 (66%), Gaps = 2/224 (0%)

30 Query: 8 IKDILWFIIPSLFGVLLLMTPFKYNGMTTVAVSVISKTNQWINAVFPIHYIILLIIFIS 67  
+KD LWF+IPS+ GV L M P + + T+ V+ ++K + ++ P I+L I +  
Sbjct: 19 LKDYLWFLIPSIIGVGLFMVPIQKDNATIPVAFLAKQLQGALDDHLPAILTIMLAIVV- 77

35 Query: 68 CVLALCYRLFRPSFIEKNDLLKEISDITIFWLIIRLIGLALGLMTVLHIGPEMVWGKETG 127  
VL+ LF+P+ KN LLK + I WL++R++G MT+L +GPE VW + TG  
Sbjct: 78 -VLSCVATLFKPNLFMKNGLLLKSLFVIHPMWLVVRVLGFIFAFMTLLQLGPEAVWSEGTG 136

40 Query: 128 GLILFDLIGGLFTIFLAAGFILPFLTEFGLLEFVGVLTPIMRPFQLPGRSAVNCVASF 187  
L+L+DL+ LFTIFL AG LPFL FGLLE GV L MRP F LPGRS+++C+AS+  
Sbjct: 137 ALLLYDLLPLFTIFLFAGLFLPFLNFGLLLEFGLVLLNKFMRPVFTLPGRSSIDCLASW 196

45 Query: 188 VGDGTIGIALTDKQYVEGYYSREAAITSTTFSVAVSITFCLXXL 231  
+GDGTIG+ LT+KQY EG+YT REAA ISTTFS VSITF + L  
Sbjct: 197 MGDGTIGVLLTNKQYEEGFYTQREAAVISTTFSVVSITFSIVVL 240

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1510

- 50 A DNA sequence (GBSx1599) was identified in *S.agalactiae* <SEQ ID 4643> which encodes the amino acid sequence <SEQ ID 4644>. This protein is predicted to be Na/H antiporter homolog (kefB). Analysis of this protein sequence reveals the following:

-1670-

Possible site: 17

&gt;&gt;&gt; Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -10.14 Transmembrane 176 - 192 ( 171 - 203)  
 INTEGRAL Likelihood = -9.34 Transmembrane 353 - 369 ( 348 - 373)  
 INTEGRAL Likelihood = -9.24 Transmembrane 3 - 19 ( 1 - 26)  
 INTEGRAL Likelihood = -7.17 Transmembrane 145 - 161 ( 142 - 168)  
 INTEGRAL Likelihood = -7.01 Transmembrane 86 - 102 ( 81 - 108)  
 INTEGRAL Likelihood = -6.53 Transmembrane 52 - 68 ( 51 - 72)  
 INTEGRAL Likelihood = -5.79 Transmembrane 24 - 40 ( 23 - 49)  
 INTEGRAL Likelihood = -5.52 Transmembrane 214 - 230 ( 209 - 233)  
 INTEGRAL Likelihood = -4.04 Transmembrane 260 - 276 ( 258 - 278)  
 INTEGRAL Likelihood = -3.66 Transmembrane 287 - 303 ( 287 - 308)  
 INTEGRAL Likelihood = -2.71 Transmembrane 113 - 129 ( 112 - 129)  
 INTEGRAL Likelihood = -2.66 Transmembrane 332 - 348 ( 330 - 349)

----- Final Results -----

bacterial membrane --- Certainty=0.5055(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAA51756 GB:X73329 Na/H antiporter homolog [Lactococcus lactis]

Identities = 208/376 (55%), Positives = 285/376 (75%), Gaps = 3/376 (0%)

Query: 1 MHIIIQITIIILLASVLATLISKRIGIPAVVGQLLVGIIGPAMLGVLHQNQVLHVLSEIG 60  
 M+ I+Q+TI+L+AS++ATL S+R+ IPAV+GQ+LVGI+I P++LGLVH VL V+SEIG  
 Sbjct: 1 MNDILQLTIVLIASLIATLASRRLKIPAVIGQMLVGIILIPSVLGLVHSGHVLVEMSEIG 60

Query: 61 VILLMFLAGLEANFDLLKKYLKPSLLVAITGVIVPMALFYFLTRLFGFQINTAIFYGLVF 120  
 VILLMFLAGLE++ +LKK K S+LVAI GVIVP+ +F + FG+ ++T+ FYG+VF  
 Sbjct: 61 VILLMFLAGLESDLTVLKKNFKASMLVAIGGVIVPLIVFGLVAFSFGYGMSTSFYFIVF 120

Query: 121 AATSISITVEVLQEYNRVKTDTGAILGAADVAVLLSVFIA--TNGSSSNIGLQI 178  
 AATS+SITVEVLQEY ++ T G+IILGAAV DD+LAVL+LS+F + GS +++ Q  
 Sbjct: 121 AATSVSITVEVLQEYKGLSTRAGSIILGAADVDDILAVLILSIFTSFKNGSGTHLFFQF 180

Query: 179 IIQLLFFVFLFICMKYLPALFKLIEKVHFFFEKYTILAILICFSLSLADKVGMSIIGS 238  
 +++LLFF FLF+ K L+P +K ++K+ K TI+A++IC LS+LAD VGMS++IGS  
 Sbjct: 181 LLELLFFAFPLFVVHK-LIPRFWKVQKLPKANKNTIVALIICLGLSLADSVGMSAVIGS 239

Query: 239 FFAGLAIGQTSFVDKVEHKISLLSYTFFIPIFFASIALPLKFDGMMSHLHTLIFTALAV 298  
 FFAGLAI QT K+E S + Y FIP+FF IA+ ++FD ++ H IL+FT LA+  
 Sbjct: 240 FFAGLAISQTEVSHKIEEYTSAGYVIFIPVFFVLIASVQFDSLHHHPWIIILLFTLLAI 299

Query: 299 LSKLIPGYFVGRGFNFESKLESITIGGGMVSRGEMALIIVQVGLAAKIISSSTYSELVIVV 358  
 L+K IP YFVG+ S ES+ IG GM+SRGEMALI+ Q+GL + II+ YSELVIV+  
 Sbjct: 300 LTKFIPAYFVGKSNKLSTGESMLIGTGMISRGEMALIVAQIGLTSAITTDEVYSELVIVI 359

Query: 359 ILSTIIAPFILKYSFK 374

IL+T++APF++K K

Sbjct: 360 ILATVLAPFLIKLVLK 375

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for  
 vaccines or diagnostics.

**Example 1511**

A DNA sequence (GBSx1600) was identified in *Sagalactiae* <SEQ ID 4645> which encodes the amino  
 acid sequence <SEQ ID 4646>. Analysis of this protein sequence reveals the following:

Possible site: 22

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

-1671-

## ----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14269 GB:Z99116 ypuA [Bacillus subtilis]

Identities = 86/319 (26%), Positives = 147/319 (45%), Gaps = 34/319 (10%)

Query: 3 IKKLLFAGLAFILFTLASPAYAASDVQKVIDETIVQPDYVLGYSLNQEORQATLQLLNVD 62  
 +KK+ LA + L P + +D + + V LG L++ + + L +N  
 Sbjct: 1 MKKIWIGMLAAAVLLLMPKVSLADA--AVGDVIV---TLGADLSESDKQKVLDEMNPV 54

Query: 63 ESRDTKVKTLTNTSSYAKIMNIADDASIQLY---SSVKIKKLGSNDTLAVNIVTPENITK 118  
 ++ T V N + + +A I SS+ I K GS +N+ T NI+  
 Sbjct: 55 DNATT-VTVTNKEEHEYLGYISNAQIGSRAISSSSITIAKKGS---LNVET-HNISG 108

Query: 119 VTEDMYRNAAVTLGIEHATISVAAPIKVTGESALAGIYYSL-KNGASVSSSENKQLAQEE 177  
 +T++MY NA +T G++ A + V AP +V+G +AL G+ + E + ++S + KQ+A +E  
 Sbjct: 109 ITDEMYLNALMTAGVKDAKVYVTAPFEVSGTAALTGLIKAYEVSSDEAISEDVKQVANQE 168

Query: 178 LSTLSGINAENKKEGYDADKLNVALTDIKSAVAKGSDLSKDDIRKIVEETLKNYHLDN 237  
 L T S + + G E A + IK AK G +K DI K V++ + L+  
 Sbjct: 169 LVTTEL-GDKIGNENAAA-----LIAKIKEEFKNGVDPNKAIDKQVDDAASD--LNV 220

Query: 238 AVTENQINLIVNFAVNLSQSNVIKNSDFTNTLNNLKDNIIVSKAGSKFKNINVNFNANKAV 297  
 +T++Q N +V S N +KN+D + D + KA K + +  
 Sbjct: 221 TLTDSDQNQLV-----SLFNKMNADI--DWGQVSDQL-DKAKDKITKFIESDEGKNFI 271

Query: 298 ESGKGFLANIWQQIVNFFQ 316  
 + F +IW IV+ F+  
 Sbjct: 272 QKVIDFFVSIWNAIVSIFK 290

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1512**

A repeated DNA sequence (GBSx1602) was identified in *S.agalactiae* <SEQ ID 4647> which encodes the amino acid sequence <SEQ ID 4648>. Analysis of this protein sequence reveals the following:

Possible site: 19  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0603(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15719 GB:Z99122 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 76/138 (55%), Positives = 91/138 (65%), Gaps = 12/138 (8%)

Query: 1 MKLKAVHHIAIIVSDYEKSKDFYVKNLGFRIIENHRPERHDYKLDLRC-GDIELEIFGN 59  
 M LK++HHIAII SDYEKSK FYV+KLGF++I+E +R ER YKLDL G +E+F  
 Sbjct: 1 MLLKSIHHIAIICSDEYKSAFYVHKLGFQVIQETREERGSYKLDLSLNGSYVIELF-- 58

Query: 60 RLDDPEYETPPQIRPNWPREACGLRHAFYVPDVEAYKVELENLGI FVEPIRYDDYTG 119  
 + PP+R RP EA GLRHAF V ++ EL GI EPIR D TG  
 Sbjct: 59 -----SFPDPPERQTRP-----EAAGLRHLAFTVGS LDKAVQELHEKGIETETPIRTDPLTG 109

-1672-

Query: 120 KKMTFFFDGDLPLELHE 137  
 K+ TFFFDGDL LPLEL+E  
 Sbjct: 110 KRFTFFFDGDLPLELYE 127

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4649> which encodes the amino acid sequence <SEQ ID 4650>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.1205(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/137 (72%), Positives = 116/137 (84%)

Query: 1 MKLKAVHHIAIIVSDYKSKDFYVKNLGFEEIRENHRPERHDYKLDLRCGDIELEIFGNR 60  
 MKL A+HH+AIIVSDY SKDFYVKNLGFEEIREN+RP++HDYKLDL CG IELEIFG  
 20 Sbjct: 2 MKLNAIHHVAIIVSDYHLSKDFYVKNLGFEEIRENYRDPKHDKLDLSCGRIELEIFGKV 61

Query: 61 LDDPEYETPPQIRGNWPREACGLRHLAFYVPDVEAYKVELENLGFVEPIRYDDYTGK 120  
 DP Y+ PP+R+ P + EACGLRHLAF V ++E+Y +L++LGI VEP+IR+DDYTG+  
 25 Sbjct: 62 TSDPNYQAPPKRVSSEPEFKSEACGLRHLAFRVNIESYVDDLKSLGIPVEPIRHDDYTGE 121

Query: 121 KMTFFFDGDLPLELHE 137  
 KMTFFFDGDLPLELHE  
 30 Sbjct: 122 KMTFFFDGDLPLELHE 138

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1513

A DNA sequence (GBSx1603) was identified in *S.agalactiae* <SEQ ID 4651> which encodes the amino acid sequence <SEQ ID 4652>. This protein is predicted to be alpha-amylase. Analysis of this protein  
 35 sequence reveals the following:

Possible site: 40  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood =-11.62 Transmembrane 14 - 30 ( 7 - 36)

40 ----- Final Results -----  
 bacterial membrane --- Certainty=0.5649(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG41778 GB:AF213261 sortase [Streptococcus gordonii]  
 Identities = 136/247 (55%), Positives = 174/247 (70%), Gaps = 2/247 (0%)

Query: 2 RNKKKSHGFFNFVRWLLVLLIIIVGLALVFNKPIRNAFIAHQSNHYQISRVSKKTIEKNK 61  
 R KK N + +L V+L++V LAL+FN IRN + +N YQ+S+VSKK IEKNK  
 50 Sbjct: 6 RRAKKKRSRRNIILNLSVILLVALALIFNSSIRNMIMVWHTNKYQVSKVSKKEIEKNK 65

Query: 62 KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPIFKGLGNTLSYGAG 121  
 SK S++F V+ +STE++L+AQ K+ LPVIGGIAIP++ +NLPIF GL N L YGAG  
 55 Sbjct: 66 ASKGSFNFKEVEPLSTEAVLNQWKAQQLPVIGGIAIPELSLNLPFNGLNAGLYYGAG 125

Query: 122 TMKENQIMGPNNYALASHHVFLTGSSKMLFSPLEHAKGKMYLTDKSKVYTYTITEI 181  
 TMKE Q M G NYALASHHVFG+TG+++MLFSPL+ AK GMK+YLTDK KVYTY+IT +

-1673-

Sbjct: 126 TMKETQEM-GKGNALASHHVFGITGANEMLFSPDRAGMKIYLTDKKVVYTSITSV 184

Query: 182 SKVTPEHVEVIDD-TPGKSQTLVLTCTDPEATERIIVHAELEKTGEFSTADESILKAFSK 240  
V PE V+V+DD G +++TLVTC D AT R IV LE + + IL F+K

Sbjct: 185 ENVEPERVDVDDAADGTAEVTLVTCEDAAATSRTIVKGVLESETPYKETPKKILNYFNK 244

Query: 241 KYNQINL 247  
YNQ+ L

Sbjct: 245 SYNQMQL 251

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4653> which encodes the amino acid sequence <SEQ ID 4654>. Analysis of this protein sequence reveals the following:

Possible site: 34  
>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -8.12 Transmembrane 18 - 34 ( 13 - 38)  
INTEGRAL Likelihood = -0.32 Transmembrane 94 - 110 ( 94 - 110)

----- Final Results -----  
bacterial membrane --- Certainty=0.4248(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA73122 GB:M77279 alpha-amylase [unidentified cloning vector]  
Identities = 60/122 (49%), Positives = 85/122 (69%)

Query: 7 RRKIKSMWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNKEAKS 67  
+ K + +W L+ +L I+GLAL+FN IR+ ++ +NS Y V+K+ +KKN ++

Sbjct: 4 KEKRRGKNWLINSLLVLLFIIGLALIFNNQIRSWVQNSRSYAVSKLKPADVKKNMARET 64

Query: 68 TFDFOAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTIELYAGATMKEE 127  
TFDF +VE +STE+V++AQ + LPVIG IAIP + INLPIFKGL N L+ GAGTMKE+

Sbjct: 65 TFDSDVESLSTEAVMKAQFENKNLPVIGAIIPSVEINLPIFKGLSNVALLTGAGTMKED 124

An alignment of the GAS and GBS proteins is shown below.

Identities = 147/245 (60%), Positives = 192/245 (78%)

Query: 2 RNKKKSHGFFNFVRWLLVLLIIVGLALVFNKPIRNAFIAHQSNHYQISRVSKKTIEKNK 61  
+ K++ ++ R LL+ +L+I+GLAL+FNKPIRN IA SN YQ+++VSKK I+KKN

Sbjct: 4 KQRRRKIKSMWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNK 63

Query: 62 KSKTSYDFSSVKSISTESILSAQTKSHNLPVIGGIAIPDVEINLPIFKGLGNTLSYGAG 121  
++K+++DF +V+ +STES+L AQ + LPVIGGIAIP++ INLPIFKGLGNTL YGAG

Sbjct: 64 EAKSTFDFOAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTIELYAG 123

Query: 122 TMKENQIMGGPNNYALASHHVFGITGSSKMLFSPLEHAKGKMKVYLTDSKVYTYTITEI 181  
TMKE Q+MGG NNY+LASHH+FG+TGSS+MLFSPLE A+ GM +YLTDK K+Y Y I ++

Sbjct: 124 TMKEEQVMGGENNYSLASHHIFGITGSSQMLFSPLEAQNQMSIYLTDKKTYBYI IKDV 183

Query: 182 SKVTPEHVEVIDDTPGKSQTLVLTCTDPEATERIIVHAELEKTGEFSTADESILKAFSKK 241  
V PE V+VIDDT G ++TLVTC D EATERIIV EL+ +F A +LKAF+

Sbjct: 184 FTVAPERVDVIDDTAGLKEVTLVTC D EATERIIVKGELKTEYDFDKAPADVLKAFNHS 243

Query: 242 YNQIN 246  
YNQ++

Sbjct: 244 YNQVS 248

SEQ ID 4652 (GBS266) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 11; MW 26kDa).

GBS266-His was purified as shown in Figure 205, lane 10.

-1674-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1514

A DNA sequence (GBSx1604) was identified in *S.agalactiae* <SEQ ID 4655> which encodes the amino acid sequence <SEQ ID 4656>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4657> which encodes the amino acid sequence <SEQ ID 4658>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

20           bacterial cytoplasm --- Certainty=0.1934(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25           Identities = 711/819 (86%), Positives = 767/819 (92%)

Query: 1   MQDKNLVDVNLTSSEKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60  
          MQD+NL+DVNLTSSEKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP  
Sbjct: 1   MQDRNLIDVNLTSSEKTSFIDYAMSVIVARALPDVRDGLKPVHRRILYGMNELGVTPDKP 60

30           Query: 61   HKKSARITGDMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNFGSMDGDGAAQRY 120  
                  HKKSARITGDMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNFGSMDGDGAAQRY  
          Sbjct: 61   HKKSARITGDMGKYHPHGDSSIYEAMVRMAQWWSYRHLVDGHGNFGSMDGDGAAQRY 120

35           Query: 121   TEARMSKIALEMLRDINKNTVDFQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMATNI 180  
                  TEARMSKIALE+LRDINKNTV+FDQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMATNI  
          Sbjct: 121   TEARMSKIALEMLRDINKNTVNFQDNYDGSEREPLVLPARFPNLLVNGATGIAVGMATNI 180

40           Query: 181   PPHNLGESIDAVKLVMDNPDVTTRELMEVIPGPDFPTGALVMGRSGIHRAYETGKGSIVL 240  
                  PPHNL ESIDAVK+VM++PD TTRELMEVIPGPDFPTGALVMGRSGIHRAY+TGKGSIVL  
          Sbjct: 181   PPHNLAESIDAVKVMMEHPDCTTRELMEVIPGPDFPTGALVMGRSGIHRAYDTGKGSIVL 240

45           Query: 241   RSRTEIETTSNGKERIVVTEFPYGVNKTKVHEHIVRLAQEKRIEGITAVRDESSREGVRF 300  
                  RSRTEIETT G+ERIVVTEFPYGVNKTKVHEHIVRLAQEKR+EGITAVRDESSREGVRF  
          Sbjct: 241   RSRTEIETTQTGRERIVVTEFPYGVNKTKVHEHIVRLAQEKRLEGITAVRDESSREGVRF 300

50           Query: 301   VIEVRRASANVILNNLFKLTSLQTNFSFNMLAIEKGVPKILSLRQIIDNYIEHQKEVIV 360  
                  VIE+RR ASA VILNNLFKLTSLQTNFSFNMLAIE GVPKILSLRQIIDNYI HQKEVI+  
          Sbjct: 301   VIEIRREASATVILNNLFKLTSLQTNFSFNMLAIEENGVPKILSLRQIIDNYISHQKEVII 360

             Query: 361   RRTFQDKAKAGARAHILEGLLVALDHLDEVITIIRNSETDTIAQELMSRFELSERQSQ 420  
                  RRT+FDK KA ARAHILEGLL+ALDHLDEVI IIRNSETD IAQ ELMSRF+LSERQSQ  
             Sbjct: 361   RRTRFDKDKAEARAHILEGLLIALDHLDEVIATIRNSETDVIAQTELMSRFDLSERQSQ 420

55           Query: 421   ILDMRLRLTGLERDKIQSEYNDLLALIALDLADILAKPERVVTIIEEMDEVKRYADAR 480  
                  ILDMRLRLTGLERDKIQSEY+DLLALIALDLADILAKPER++TIIKEEMDE+KRYA+ R  
             Sbjct: 421   ILDMRLRLTGLERDKIQSEYDDLALIALDLADILAKPERITTIIEEMDEIKRYANPR 480

60           Query: 481   RTELMIGEVLSEDEDLIEEDVLITLSNKGYIKRLAQDEFRAQKRGGRGIQGTGVNND 540  
                  RTELM+GEVLSEDEDLIEEDVLITLSNKGYIKRLAQDEFRAQKRGGRG+QGTGVNND

-1675-

Sbjct: 481 RTEL MVGEVLSLEDEDLIEEDVLITLSNKG YIKRLAQDEFRAQKRGGRGVQGTGVNND 540

Query: 541 FVREL VSTSTHDTVLFFTNLGRVYRLKAYEIP EYGR TAKGLPIVNL LKLEGETIQTIIN 600  
FVREL+STSTHDT+LFFTN GRVYRLKAYEIP EYGR TAKGLPIVNL LK++GETIQTIIN

5 Sbjct: 541 FVRELISTHDTLLFFTNFGRVYRLKAYEIP EYGR TAKGLPIVNL LKLEGETIQTIIN 600

Query: 601 ARKEDVANKYFFFTTQQGIVKRTSVSEFSNIRQNGLR AINLKENDELINVL LIDENEDVI 660  
ARKE+ A K FFFTT+QGIVKRT VSEF+NIRQNGLR A+ LKE D+LINVLL +D+I

10 Sbjct: 601 ARKEETAGKSFFFTTKQGIVKRT EVSEFNNIRQNGLR ALKLKEGDQLINVL LITSGQDDII 660

Query: 661 IGTRTGYSVRFKVNAVRNMGR TATGVRGVNLREGDKVVGASRI VNGQEVLIITEKG YGKR 720  
IGT +GYSVRF ++RNMGR+ATGVRGV LRE D+VVGASRI + QEVL+ITE G+GKR

Sbjct: 661 IGTHSGYSVRFEASIRNMGR SATGVRGVKLREDDRVVGASRI QDNQEVLVITENGFGKR 720

15 Query: 721 TEASEYPTKGRGGKG IKTANITAKNGPLARLVTINGNEDIMVITDTGV IIRTNVANISQT 780  
T A++YPTKGRGGKG IKTANIT KNG LA LVT++G EDIMVIT+ GVIIRTNVANISQT

Sbjct: 721 TSATDYPTKGRGGKG IKTANITPKNGQLAGLVTVDGTEDIMVITNKGVIIRTNVANISQT 780

20 Query: 781 GRSTMGVKVMRLDQEAKIVT VALVEQEIEDKSNIEDTKE 819  
GR+T+GVK+M+LD +AKIVT LV+ E + I +E

Sbjct: 781 GRATLGVKIMKLDADAKIVTFTLVQPEDSSIAEINTDRE 819

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 25 Example 1515

A DNA sequence (GBSx1605) was identified in *S. agalactiae* <SEQ ID 4659> which encodes the amino acid sequence <SEQ ID 4660>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA04010 GB:AJ000336 L-lactate dehydrogenase [*Streptococcus pneumoniae*]  
Identities = 290/329 (88%), Positives = 313/329 (94%), Gaps = 1/329 (0%)

40 Query: 1 MTATKQHKK VILVGDGAVGSSYAFALVNQGI AQELGIIEIPALFDKAVGDAEDLSHALAF 60  
MT+TKQHKK VILVGDGAVGSSYAFALVNQGI AQELGIIEIP L +KAVGDA DLSHALAF

Sbjct: 1 MTSTKQHKK VILVGDGAVGSSYAFALVNQGI AQELGIIEIPQLHEKAVGDALDLSHALAF 60

45 Query: 61 TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120  
TSPKKIYAA Y+DCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGF GI

Sbjct: 61 TSPKKIYAAQYSDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFKGI 120

50 Query: 121 FLVAANPVDVLTYSTWKFSGF PKERVIGSGTSLDSARFRQALADKIGVDARSVHAYIMGE 180  
FLVAANPVDVLTYSTWKFSGF PKERVIGSGTSLDSARFRQALA+K+ VDARSVHAYIMGE

Sbjct: 121 FLVAANPVDVLTYSTWKFSGF PKERVIGSGTSLDSARFRQALAEKLDVDARSVHAYIMGE 180

Query: 181 HGDSEFAVWSHANVAGVQLEQLQENRDIDEQGLVDL FISRDAAYS IINKKGATYYGIA 240  
HGDSEFAVWSHAN+AGV LE++L++ +++ E L++LF VRDAAY+IINKKGATYYGIA

55 Sbjct: 181 HGDSEFAVWSHANTAGVNLEEFKDTQNVQEAELIELFEGVRDAAYTIINKKGATYYGIA 240

Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDV KDVFIGQPAIVGAHGIVRPVNIPLNDAEL 300  
VALARITKAILDDENAVLPLSV+QEGQYG V++VFIGQPA+VGAHGIVRPVNIPLNDAE

Sbjct: 241 VALARITKAILDDENAVLPLSVFQEGQYG- VENVFIGQPAVGAHGIVRPVNIPLNDAET 299

60 Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329  
QKMQASA++L+ IDEAWKNPEFQEASKN

Sbjct: 300 QKMQASAKELQAIIDEAWKNPEFQEASKN 328

-1676-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4661> which encodes the amino acid sequence <SEQ ID 4662>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -1.17 Transmembrane 106 - 122 ( 106 - 122)

----- Final Results -----

bacterial membrane --- Certainty=0.1468(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB81558 GB:U60997 L(+)-lactate dehydrogenase [Streptococcus  
 bovis]  
 Identities = 278/329 (84%), Positives = 297/329 (89%), Gaps = 2/329 (0%)

Query: 1 MTATKQHKKVLVGDGAVGSSYAFALVTQNIQAELGIIDIFK--EKTQGDADLSHALAF 58  
 MTATKQHKKVLVGDGAVGSSYAFALV Q IAQELGII+I + K GDAEDLSHALAF  
 Sbjct: 1 MTATKQHKKVLVGDGAVGSSYAFALVNQGIQAELGII+I+PQLFNKAVGDAEDLSHALAF 60

Query: 59 TSPKKIYAADYSDCHDADLVVLTAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118  
 TSPKKIYAA Y DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VT++V SGFKGI  
 Sbjct: 61 TSPKKIYAAKYEDCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTEVVKSGFKGI 120

Query: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSHAYIMGE 178  
 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA K+ VDARSHAYIMGE  
 Sbjct: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAEKLDVDARSHAYIMGE 180

Query: 179 HGDSEFAVWSHANVAGVGLYDQLQNRDIDEQGLVDLFI SVRDAAYSIIINKKGATFYGIA 238  
 HGDSEFAVWSHANVAGV L +L+ ++++E LV+LF VRDAAYSIIINKKGATFYGIA  
 Sbjct: 181 HGDSEFAVWSHANVAGVNLESYLDVQNVEEAELVELFEGVRDAAYSIIINKKGATFYGIA 240

Query: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298  
 VALARITKAIL+DENAVLPLSVFQEGQY V DCYIGQPAIVGA+GIVRPVNIPLNDAE  
 Sbjct: 241 VALARITKAILNDENAVLPLSVFQEGQYANVTDCYIGQPAIVGAHGIVRPVNIPLNDAEQ 300

Query: 299 QKMQASANQLKAIIDEAFKEEFASAAKN 327  
 QKM+ASA +LKAIIDEAF+KEEFASA KN  
 Sbjct: 301 QKMEASAKELKAIIDEAFSKEEFASACKN 329

An alignment of the GAS and GBS proteins is shown below.

Identities = 286/329 (86%), Positives = 299/329 (89%), Gaps = 2/329 (0%)

Query: 1 MTATKQHKKVLVGDGAVGSSYAFALVNQGIQAELGII+I+PALFDKAVGDAEDLSHALAF 60  
 MTATKQHKKVLVGDGAVGSSYAFALV Q IAQELGII+I +K GDAEDLSHALAF  
 Sbjct: 1 MTATKQHKKVLVGDGAVGSSYAFALVTQNIQAELGIIDI--FKEKTQGDADLSHALAF 58

Query: 61 TSPKKIYAATYADCADADLVVITAGAPQKPGETRLDLVGKNLAINKSIVTQVVESGFNGI 120  
 TSPKKIYAA Y+DC DADLVV+TAGAPQKPGETRLDLV KNL INK +VTQ+V SGF GI  
 Sbjct: 59 TSPKKIYAADYSDCHDADLVVLTAGAPQKPGETRLDLVEKNLRINKEVVTQIVASGFKGI 118

Query: 121 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALADKIGVDARSHAYIMGE 180  
 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALA KIGVDARSHAYIMGE  
 Sbjct: 119 FLVAANPVDVLTYSTWKFSGFPKERVIGSGTSLDSARFRQALAAKIGVDARSHAYIMGE 178

Query: 181 HGDSEFAVWSHANVAGVQLQWLQENRDIDEQGLVDLFI SVRDAAYSIIINKKGATYYGIA 240  
 HGDSEFAVWSHANVAGV L WLQ NRDIDEQGLVDLFI SVRDAAYSIIINKKGAT+YGIA  
 Sbjct: 179 HGDSEFAVWSHANVAGVGLYDQLQNRDIDEQGLVDLFI SVRDAAYSIIINKKGATFYGIA 238

Query: 241 VALARITKAILDDENAVLPLSVYQEGQYGDVKDVFQIPQPAIVGAHGIVRPVNIPLNDAEL 300  
 VALARITKAILDDENAVLPLSV+QEGQY V+D +IGQPAIVGA+GIVRPVNIPLNDAEL  
 Sbjct: 239 VALARITKAILDDENAVLPLSVFQEGQYEGVEDCYIGQPAIVGAYGIVRPVNIPLNDAEL 298

-1677-

Query: 301 QKMQASAEQLKDIIDEAWKNPEFQEASKN 329  
 QKMQASA QLK IIDEA+ EF A+KN  
 Sbjct: 299 QKMQASANQLKAIIDEAFAKEEFASAANK 327

5

SEQ ID 4660 (GBS312) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 7; MW 40kDa).

GBS312-His was purified as shown in Figure 205, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10

### Example 1516

A DNA sequence (GBSx1606) was identified in *S.agalactiae* <SEQ ID 4663> which encodes the amino acid sequence <SEQ ID 4664>. This protein is predicted to be NADH oxidase (nox). Analysis of this protein sequence reveals the following:

15

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20

bacterial cytoplasm --- Certainty=0.1888 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25

>GP:AAC26485 GB:AF014458 NADH oxidase [Streptococcus pneumoniae]  
 (ver 2)  
 Identities = 363/458 (79%), Positives = 408/458 (88%), Gaps = 3/458 (0%)

30

Query: 1 MSKIVVVGTHAGTAAIKTMLSNEYGEANEIVTFDQNSNISFLGCGMALWIGEQIDGPEGL 60  
 MSKIVVVG NHAGTA I TML N+G NEIV FDQNSNISFLGCGMALWIGEQIDG EGL  
 Sbjct: 1 MSKIVVVGANHAGTACINTMLDNFGNENEIVVFDQNSNISFLGCGMALWIGEQIDGA EGL 60

35

Query: 61 FYSDKEQLES MGAKVYMNSPVLNIDYDKKEVTALVDGKEHVESYEKLI LILATGSQPIIPPI 120  
 FYSDKE+LE+ GAKVYMNSPVL+IDYD K VTA V+GKEH ESYEKLI ATGS PI+PPI  
 Sbjct: 61 FYSDKEKLEAKGAKVYMNSPVLSDIDYDNKVVTA EVEGKEHKESEYKLI FATGSTPILPPI 120

40

Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSEEVIEKLAKPG--INRVA VVGAGYIGVELAEA 178  
 +GVEI +G+REFKATLEN+QFVKLYQN+EEVI KL+ ++R+AVVG .GYIGVELAEA  
 Sbjct: 121 EGVEIVKGNREFKATLENVQFVKLYQNAEEVINKLSDKSQHLDR IAVVGGGYIGVELAEA 180

45

Query: 179 FQRIGKEVTLV DVADTCMGYYDRDFTDMMSKNLEDHGIRLAFGQAVQAVEGDGKVERLV 238  
 F+R+GKEV LVD+ DT + GYYD+DFT MM+KNLEDH IRLA GQ V+A+EGDGKVERL+  
 Sbjct: 181 FERLGKEVVLVDIVDTVLNGYYDKDFTQMMAKNLEDHNIRLALGQTVKAIEGDGKVERLI 240

Query: 239 TDKETFDVDMVILAVGFRPNTEL GAGKLDTFRNGAWVVDKKQETS VKDVYAIGDCATIWD 298  
 TDKE+FDVDMVILAVGFRPNT L GK++ FRNGA++VDKKQETS+ VYA+GDCAT++D  
 Sbjct: 241 TDKESFDVDMVILAVGFRPNTALADGKIELFRNGAFLVDKKQETSIPGVYAVGDCATVYD 300

50

Query: 299 NSRDDINYLALASNAVRTGIVAAHNACGTELEGAGVQGSNGISITYGLNMVSTGLTLEKAK 358  
 N+R D +YIALASNAVRTGIV A+NACG ELEG GVQGSNGISITYGL+MVSTGLTLEKAK  
 Sbjct: 301 NARKDTSYIALASNAVRTGIVGAYNACGHELEGIGVQGSNGISITYGLHVMSTGLTLEKAK 360

55

Query: 359 QAGYNVETG FNDLQKPEFIKHNNHEVAIKIVYDKDSRVILGCQMVSHS-DVSMGIHMF 417  
 AGYNA ETGFNDLQKPEF+KH+NHEVAIKIV+DKDSR ILG QMVSH+ +SMGIHMF  
 Sbjct: 361 AAGYNATETGFNDLQKPEFMKHDNHEVAIKIVFDKDSREILGAQMVSHDIAISMGIHMF 420

Query: 418 LAIQEKVTIEKLALTDIFFLPHFNKPNYITMAALGAK 455  
 LAIQE VTI+KLALTD+FFLPHFNKPNYITMAAL A+  
 Sbjct: 421 LAIQEHVTIDKLALTDLFFLPHFNKPNYITMAALTAE 458

-1678-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4665> which encodes the amino acid sequence <SEQ ID 4666>. Analysis of this protein sequence reveals the following:

Possible site: 14

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2068(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 362/456 (79%), Positives = 403/456 (87%)

Query: 1 MSKIVVVGTHAGTAAIKTMSNYGEANEIVTFDQNSNISFLGCGMALWIGEIDGPEGL 60  
MSKIVVVG NHAGTA IKTML+NYG+ANEIV FDQNSNISFLGCGMALWIGEIQI GPEGL  
Sbjct: 1 MSKIVVVGANHAGTACIKTMLTNYGDANEIVVFDQNSNISFLGCGMALWIGEIQIAGPEGL 60

Query: 61 FYSKDKEQLES MGAKVYMNSPVLNIDYDKKEVTALVDGKEHVESYEKLILATGSQPIIPPI 120  
FYSKDKE+LES+GAKVYM SPV +IDYD K VTALVDGK HVE+Y+KLI ATGSQPI+PPI  
Sbjct: 61 FYSKDKEELES LGAKVYMESPVQSIDYDAKTVTALVDGKNHVETYDKLIFATGSQPIIPPI 120

Query: 121 KGVEIQEGSREFKATLENLQFVKLYQNSSEEVIEKLAKPGINRVAVVGAGYIGVELAEAFQ 180  
KG EI+EGS EF+ATLENLQFVKLYQNS +VI KL I RVAVVGAGYIGVELAEAFQ  
Sbjct: 121 KGABIEKESLEFEATLENLQFVKLYQNSADVIKLENKDIKRVAVVGAGYIGVELAEAFQ 180

Query: 181 RIGKEVTLVDVADTCMGYYDRDFTDMSKNLEDHGIRLAFGQAVQAVEGDGKVERLVT 240  
R GKEV L+DV DTC+ GYYDRD TD+M+KN+E+HGI+LAFG+ V+ V G+GKVE+++TD  
Sbjct: 181 RKGKEVVLIDVVDTCLAGYYDRDLTDLMAKNMEEHGIQLAFGETVKEVAGNGKVEKIITD 240

Query: 241 KETFDVDMVILAVGFRPNTLGGAGKLDTFRNGAWVVDKQETSVKDVYAIGDCATIWDNS 300  
K +DVDVDMVILAVGFRPNT LG GK+D FRNGA++V+K+QETS+ VYAIGDCATI+DN+  
Sbjct: 241 KNEYDVDMVILAVGFRPNTTLGNGKIDLFRNGAFLVNKRQETSIPGVYAIGDCATIYDNA 300

Query: 301 RDDINYIALASNAVRTGIVAHAHACGTELEGAGVQGSNGISITGLNMVSTGLTLEKAKQA 360  
D NYIALASNAVRTGIVAHAHACGT+LEG GVQGSNGISITGL+MVSTGLTLEKAK+  
Sbjct: 301 TRDTNYIALASNAVRTGIVAHAHACGTDLLEGIGVQGSNGISITGLHVMVSTGLTLEKAKRL 360

Query: 361 GYNAVETGFNDLQKPEFIKHNHEVAIKIVYDKDSRVILGCQMVSHEDVSMGIHMFSLAI 420  
G++A T + D QKPEFI+H N V IKIVYDKDSR ILG QM + EDVSMGIHMFSLAI  
Sbjct: 361 GFDAAVTEYTDNQKPEFIEHGNFPVTIKIVYDKDSRILGAQMAAREDVSMGIHMFSLAI 420

Query: 421 QEKVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456

QE VTIEKLALTDIFFLPHFNKPNYITMAALGAKD

Sbjct: 421 QEGVTIEKLALTDIFFLPHFNKPNYITMAALGAKD 456

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1517

A DNA sequence (GBSx1607) was identified in *S.agalactiae* <SEQ ID 4667> which encodes the amino acid sequence <SEQ ID 4668>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2319(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-1679-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1518

- 5 A DNA sequence (GBSx1608) was identified in *S.agalactiae* <SEQ ID 4669> which encodes the amino acid sequence <SEQ ID 4670>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

10	INTEGRAL	Likelihood = -7.75	Transmembrane	160 - 176 ( 157 - 179)
	INTEGRAL	Likelihood = -7.38	Transmembrane	73 - 89 ( 70 - 97)
	INTEGRAL	Likelihood = -5.47	Transmembrane	289 - 305 ( 284 - 312)
	INTEGRAL	Likelihood = -4.09	Transmembrane	107 - 123 ( 106 - 124)
	INTEGRAL	Likelihood = -3.24	Transmembrane	43 - 59 ( 43 - 59)
	INTEGRAL	Likelihood = -1.91	Transmembrane	258 - 274 ( 258 - 275)
15	INTEGRAL	Likelihood = -1.33	Transmembrane	234 - 250 ( 233 - 251)
	INTEGRAL	Likelihood = -0.00	Transmembrane	209 - 225 ( 209 - 225)

----- Final Results -----

20	bacterial membrane --- Certainty=0.4100(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9805> which encodes amino acid sequence <SEQ ID 9806> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 172/318 (54%), Positives = 234/318 (73%)

30	Query: 5	LSLTTFIFALLFSSMLIYATPLIFTSIGGTFSEGGIVNVGLEGIMVIGAFSGVVFNLFEFA 64
		+ + I +++ + L+YA PLI T++GG FSER G+VN+GLEG+M+IGAF+ V+FNL F
	Sbjct: 1	MDIVQILSIIVPATLVYAAPLILTALGGVFSERSGVVNIGLEGLMIIGAFTSVLFNLFFG 60
35	Query: 65	SVFGDATPWISVLVGLVGLIFSVIHAVATVNFADHIISGTVLNLMAVFLIKVLY 124
		G A PW+S+L G +FS+IHA A ++FRAD +SG +N++A +F++K++Y
	Sbjct: 61	QELGAAAPWLSLLAAMAAGALFSLIHAAAISFRADQTVSGVAINMLALGATLFIKLIY 120
40	Query: 125	NKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYIAILFSLAWFIFYKTRFGLR 184
		K QTD I E F K P L DIP +G IFF +AI +F++WFIL+KT FGLR
	Sbjct: 121	GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYSILAIALAFISWFILKTPFGLR 180
45	Query: 185	LRVGEHPQAADTLGINVYLMRYSGVLISGFLGGIGGAVYAQSISVNFAATTILGPGFIS 244
		+RSVGEHP AADT+GINVY MRY GV+ISG GG+GG VYA +I+++F +TI G GFI+
	Sbjct: 181	IRSVGEHPMAADTMGINVYKMYIGVMISGLFGGLGGVYASTIALDFTHSTISGQGFIA 240
50	Query: 245	LAAMIFGKWNPIGAMLASLFFGLSQSLAVIGSHPLLSNIPTVYLQIAPYVLTIIIVLAAF 304
		LAA++FGKW+PIGA+ A+LFFG +QSL++IGS LPL +IP VY+ +APY+LTI+ L F
	Sbjct: 241	LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLPLPKDIPNVYMLMAPYILTILALTGF 300
55	Query: 305	FGQAVAPKADGINYIKTK 322
		G+A APKA+G+ YIK K
	Sbjct: 301	IGRADAPKANGVPYIKGK 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4671> which encodes the amino acid sequence <SEQ ID 4672>. Analysis of this protein sequence reveals the following:

Possible site: 22

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.92	Transmembrane	73 - 89 ( 69 - 97)
----------	--------------------	---------------	--------------------

-1680-

INTEGRAL Likelihood = -5.04 Transmembrane 160 - 176 ( 158 - 177)  
 INTEGRAL Likelihood = -4.62 Transmembrane 289 - 305 ( 284 - 312)  
 INTEGRAL Likelihood = -3.98 Transmembrane 234 - 250 ( 232 - 251)  
 INTEGRAL Likelihood = -2.13 Transmembrane 107 - 123 ( 106 - 123)  
 INTEGRAL Likelihood = -2.02 Transmembrane 43 - 59 ( 43 - 59)  
 INTEGRAL Likelihood = -0.53 Transmembrane 258 - 274 ( 258 - 274)

----- Final Results -----

bacterial membrane --- Certainty=0.4567(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB15146 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]  
 Identities = 176/318 (55%), Positives = 239/318 (74%)

Query: 5 MSLVTIFALLMSSMLIYATPLIFTSIGGTFSESGVVNVGLEGIMVMGAFSGIVFNLEFA 64  
 M +V I +++++ L+YA PLI T++GG FSESGVVN+GLEG+M++GAF+ ++FNL F  
 Sbjct: 1 MDIVQILSIIVPATLVYAAPLIITALGGVFSESGVVNIGLEGLMIIGAFSTVLFNLF 60

Query: 65 ETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLAPSFVFLVKAMY 124  
 + G A PW+++L G +FSLIHA A I+FRAD VSG +N+LA +F+VK +Y  
 Sbjct: 61 QELGAAAPWLSLLAAMAAGALFSLIHAAAAISFRADQTVSGVAINMLALGATLFIVKLIY 120

Query: 125 GKGQTDNIQQSFGKFDFFGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTRFGLR 184  
 GK QTD I + F K PGL IPV+G IFF + AIA +F +WF+L+KT FGLR  
 Sbjct: 121 GKAQTDKIPEPFYKTKIPGLGDIPVLGKIFFSDVYYTISILAIALAFISWFLEFKTPFGLR 180

Query: 185 LRSVGEHPQAADTLGINVYLMKYGVGMISGFLGGIGGAVYAQSISVNFAVTTILGPGFIA 244  
 +RSVGEHP AADT+GINVY M+Y GVMISG GG+GG VYA +I+++F +TI G GFIA  
 Sbjct: 181 IRSVGEHPMAADTMGINVYKMYIGVMISGLFGGLGGVYASTIALDFTHTSISGQGFIA 240

Query: 245 LAAMIFGKWNPVGAMLSLFFGLSQSLAVIGAQPLLEKIPTVYLQIAPYMTIIILAAF 304  
 LAA++FGKW+P+GA+ ++LFFG +QSL++IG+ LPL + IP VY+ +APY++TI+ L F  
 Sbjct: 241 LAALVFGKWHPIGALGAALFFGFAQSLSIIGSLPLPKDIPNVYMLMAPYIILTILALTGF 300

Query: 305 FGQAVAPKADGINYIKSK 322  
 G+A APKA+G+ YIK K  
 Sbjct: 301 IGRADAPKANGVPYIKGK 318

An alignment of the GAS and GBS proteins is shown below.

Identities = 272/322 (84%), Positives = 301/322 (93%)

Query: 1 MVSKLSLTITIFALLFSSMLIYATPLIFTSIGGTFSESGIVNVGLEGIMVIGAFSGVVFN 60  
 +V+K+SL TIFALL SSMLIYATPLIFTSIGGTFSESG+VNVGLEGIMV+GAFSG+VFN  
 Sbjct: 1 VVNKMSLVITIFALLMSSMLIYATPLIFTSIGGTFSESGVVNVGLEGIMVMGAFSGIVFN 60

Query: 61 LEFASVFGDATPWISVLVGGVLGLIFSVIHAVATVNFADHIIISGTVLNLMAPSLAVFLI 120  
 LEFA FG ATPWI+VLVGG+VGLIFS+IHAVAT+NFRADHI+SGTVLNL+APS AVFL+  
 Sbjct: 61 LEFAETFGKATPWIAVLVGGIVGLIFSLIHAVATINFRADHIVSGTVLNLAPSFVFLV 120

Query: 121 KVLNKGQTDNIQESFGKFNFPILSDIPFVGDIFFKGTSLVGYIAIFSLAWFIFYKTR 180  
 K +Y KQTDNIQ+SFGKF+FP LS IP +GDIFFK TSL+GY AI FSF AWF+LYKTR  
 Sbjct: 121 KAMYGKGQTDNIQQSFGKFDFFGLSQIPVIGDIFFKNTSLIGYFAIAFSFFAWFLLYKTR 180

Query: 181 FGLRLRSVGEHPQAADTLGINVYLMRYSGVLISGFLGGIGGAVYAQSISVNFAATILGP 240  
 FGLRLRSVGEHPQAADTLGINVYLM+Y GV+ISGFLGGIGGAVYAQSISVNFA TTILGP  
 Sbjct: 181 FGLRLRSVGEHPQAADTLGINVYLMKYGVGMISGFLGGIGGAVYAQSISVNFAVTTILGP 240

Query: 241 GFISLAAMIFGKWNPIGAMLASLFFGLSQSLAVIGASHLPLLSNIPTVYLQIAPYVLTIIIV 300  
 GFT+LAAMIFGKWNPI+GAML+SLFFGLSQSLAVIG+ LPLL IPTVYLQIAPY++TII+  
 Sbjct: 241 GFIALAAMIFGKWNPIGAMLSLFFGLSQSLAVIGASHLPLLEKIPTVYLQIAPYMTIIII 300

Query: 301 LAAFFGQAVAPKADGINYIKTK 322  
 LAAFFGQAVAPKADGINYIK+K  
 Sbjct: 301 LAAFFGQAVAPKADGINYIKSK 322

[illegible]

-1682-

290 300 310

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 5 Example 1519

A DNA sequence (GBSx1609) was identified in *S.agalactiae* <SEQ ID 4673> which encodes the amino acid sequence <SEQ ID 4674>. This protein is predicted to be ribose/galactose ABC transporter, permease protein (rbsC-1). Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood =-14.59  Transmembrane  205 - 221 ( 200 - 228)
    INTEGRAL    Likelihood =-13.69  Transmembrane   21 -  37 (  13 -  45)
    INTEGRAL    Likelihood = -7.27  Transmembrane  302 - 318 ( 290 - 321)
    INTEGRAL    Likelihood = -7.17  Transmembrane  115 - 131 ( 111 - 138)
15  INTEGRAL    Likelihood = -4.25  Transmembrane  251 - 267 ( 250 - 268)
    INTEGRAL    Likelihood = -2.97  Transmembrane   63 -  79 (  63 -  80)
    INTEGRAL    Likelihood = -2.87  Transmembrane  333 - 349 ( 328 - 349)

----- Final Results -----
20      bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8831> which encodes amino acid sequence <SEQ ID 8832> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1  Crend: 6
SRCFLG: 0
McG: Length of UR:  24
      Peak Value of UR:  3.06
      Net Charge of CR:  3
30  McG: Discrim Score:  12.53
      GvH: Signal Score (-7.5): -5.31
      Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
35  Amino Acid Composition: calculated from 1
      ALOM program  count: 7 value: -14.59 threshold:  0.0
      INTEGRAL    Likelihood =-14.59  Transmembrane  196 - 212 ( 191 - 219)
      INTEGRAL    Likelihood =-13.69  Transmembrane   12 -  28 (   4 -  36)
      INTEGRAL    Likelihood = -7.27  Transmembrane  293 - 309 ( 281 - 312)
40  INTEGRAL    Likelihood = -7.17  Transmembrane  106 - 122 ( 102 - 129)
      INTEGRAL    Likelihood = -4.25  Transmembrane  242 - 258 ( 241 - 259)
      INTEGRAL    Likelihood = -2.97  Transmembrane   54 -  70 (  54 -  71)
      INTEGRAL    Likelihood = -2.87  Transmembrane  324 - 340 ( 319 - 340)
      PERIPHERAL  Likelihood =  0.16    133
45  modified ALOM score:  3.42
      icml HYPID: 7  CFP: 0.684

*** Reasoning Step: 3

50  ----- Final Results -----
      bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

55 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15145 GB:Z99120 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 154/349 (44%), Positives = 220/349 (62%), Gaps = 6/349 (1%)

Query: 10 MSKKAQKIAVPLISVVLGIILGAIIMLIFGYDPLWGYEGLFQTAFGSIKNIGEIFRAMGP 69

-1683-

M K+ + VPLI+++LG+ GA+IML+ GY GY L+ FG I +GE R + P  
 Sbjct: 1 MVKRLSHLLVPLIAIILGLAAGALIMLVSGYSVASGYSALWNGIFGEIYYVGETIRQITP 60  
 Query: 70 LILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHDMPRPAMILCTIIIGIVAGG 129  
 5 IL L + A R G FNIG+ GQ L GW AA W + D P + +I AGG  
 Sbjct: 61 YILSGLAVAFARTGLFNIGVEGQLLVGWTAADVWGTAF-DGPAYIHLPLALITAAAAGG 119  
 Query: 130 ITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRVFPKSIMRTSDSSVYVSA 189  
 + G IPGIL+A EVIVTIMMNYI L+ N I+ V D + + +AS  
 10 Sbjct: 120 LWGFIPGILKARFYVHEVIVTIMMNYIALHMTNYIISNVLTDH---QDKTGKIHESASL 175  
 Query: 190 QTDWLSSLTNNSRINIGIFIAIIAVVLVWFLNKTTLGFEIRSVGLNPNASEYAGMSAKR 249  
 ++ +L +T+ SR+++GI +A++A V++WF++NK+T GFE+R+VG N +AS+YAGMS ++  
 Sbjct: 176 RSPFLEQITDYSRLHLGLIIVALLAAVIMWFIINKSTKGFELRAVGFNQHASQYAGMSVRK 235  
 15 Query: 250 TIILSMIISGAFAGLGGVVEGLGTFENVFVQPSLAIGFDGMAVSLAANSPIGILFAAF 309  
 I+ SM+ISGAFAGL G +EGLGTFE V+ + +GFDG+AV+LL N+ +G++ AA  
 Sbjct: 236 NIMTSMIISGAFAGLAGAMEGLGTFEYAAVKGAFITGVGFDGIAVALLGNTAVGVVLAAC 295  
 Query: 310 LFGVLSVGAPGMNI-AGIPPELIKVVITASIIFFVGVHYIIIEYVIKPKKQ 357  
 L G L +GA M I +G+P E++ +V A II FV Y I +V+ K+  
 20 Sbjct: 296 LLGGLKIGALNMPIESGVPSEVVDIVIAIILFVASSYAIRFVMGKLLK 344

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2149> which encodes the amino acid  
 25 sequence <SEQ ID 2150>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have an uncleavable N-term. signal seq  
 30 INTEGRAL Likelihood = -12.74 Transmembrane 205 - 221 ( 200 - 228)  
 INTEGRAL Likelihood = -12.42 Transmembrane 21 - 37 ( 14 - 45)  
 INTEGRAL Likelihood = -7.22 Transmembrane 115 - 131 ( 111 - 135)  
 INTEGRAL Likelihood = -4.78 Transmembrane 251 - 267 ( 249 - 269)  
 INTEGRAL Likelihood = -2.50 Transmembrane 70 - 86 ( 69 - 86)  
 INTEGRAL Likelihood = -2.34 Transmembrane 302 - 318 ( 300 - 318)  
 35 INTEGRAL Likelihood = -1.44 Transmembrane 148 - 164 ( 147 - 165)  
 INTEGRAL Likelihood = -1.33 Transmembrane 326 - 342 ( 326 - 342)  
 ----- Final Results -----  
 bacterial membrane --- Certainty=0.6095(Affirmative) < succ>  
 40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 293/358 (81%), Positives = 333/358 (92%), Gaps = 1/358 (0%)  
 45 Query: 6 RRREMSKKAQKIAVPLISVVLGIIILGAIIMLIFGYDPLWGYEGLFQTAFGSIKNIGEIFR 65  
 RR+ MSK AQKIAVPLISV+LG +LGAIIM+IFGYDP+WGYEGLFQ AFGS+KNIGEIFR  
 Sbjct: 6 RRKVMKNAQKIAVPLISVVLGFLILGAIIMVIFGYDPIWGYEGLFQIAFGSVKNIGEIFR 65  
 Query: 66 AMGPLILIALGFSVASRAGFFNIGLPGQALSGWIAAGWFALSHDMPRPAMILCTIIIGI 125  
 50 +MGPLILIALGF+VASRAGFFN+GL GQAL+GWI+AGWFAL +PDMPRP +IL T +IG+  
 Sbjct: 66 SMGPLILIALGFTVASRAGFFNVGLSGQALAGWISAGWFALLNPDMRPRLILMTALIGM 125  
 Query: 126 VAGGITGAIPGILRAYLGTSEVIVTIMMNYIVLYSGNAIVQRVFPKSIMRTSDSSVYVSA 185  
 +AGGI GAIPGILRAYLGTSEVIVTIMMNYI+LY GNAIVQR +P+S+ ++ DS++ VS  
 55 Sbjct: 126 IAGGIAGAIPGILRAYLGTSEVIVTIMMNYIILYVGNNAIVQRGYPESVKQSIDSTIQVSD 185  
 Query: 186 NASYQTDWLSSLTNNSRINIGIFIAIIAVVLVWFLNKTTLGFEIRSVGLNPNASEYAGM 245  
 NASYQT WLS+LTNNSRINIGIF AIIA+ L+WFLNKTTLGFEIRSVGLNP+ASEYAGM  
 Sbjct: 186 NASYQTHWLSALTNNSRINIGIFIAIIAIIWFLNKTTLGFEIRSVGLNPHASEYAGM 245  
 60 Query: 246 SAKRTIILSMIISGAFAGLGGVVEGLGTFENVFVQPSLAIGFDGMAVSLAANSPIGIL 305  
 S+KRTIILSMIISGA AGLGGVVEGLGTFENVFVQ SSLA+GFDGMAVSLAANSP+GI  
 Sbjct: 246 SSKRTIILSMIISGALAGLGGVVEGLGTFENVFVQSSSLAVGFDGMAVSLAANSPGIF 305  
 Query: 306 FAAFLFGVLSVGAPGMNIAGIPPELIKVVITASIIFFVGVHYIIE-YVIKPKKQMKGGK 362

-1684-

F++FLFGVL++GAPGMNIAGIPPEL+KVVVTASIIFFVG HY+IE Y+I+PKK +KGGK  
 Sbjct: 306 FSSFLFGVLNIGAPGMNIAGIPPELVKVVVTASIIFFVGSHYLIERYIIRPKKLVKGGK 363

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 5 vaccines or diagnostics.

### Example 1520

A DNA sequence (GBSx1610) was identified in *S.agalactiae* <SEQ ID 4675> which encodes the amino  
 acid sequence <SEQ ID 4676>. This protein is predicted to be sugar ABC transporter, ATP-binding protein  
 (mgIA). Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3851(Affirmative) < succ>  
 15 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9803> which encodes amino acid sequence <SEQ ID 9804>  
 was also identified.

20 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15144 GB:Z99120 similar to ABC transporter (ATP-binding  
 protein) [Bacillus subtilis]  
 Identities = 311/497 (62%), Positives = 396/497 (79%), Gaps = 1/497 (0%)  
 25 Query: 14 VIEMKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDGGQIFI 73  
 VIEM I K F VAND+INL V+KGEIHALLGENGAGKSTLMN+L GL +P G+I +  
 Sbjct: 4 VIEMLNIRKAPPGIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRV 63  
 30 Query: 74 NGQPVTIDSPSKSSQLGIGMVHQHFMVLEAFTVAENIVLGNETTQNGVLDIKTAAKEIKE 133  
 G+ V I+SP+K++ LGIGMVHQHFMV+ FTVAENI+LG E + G +D K A +E+++  
 Sbjct: 64 RGEKVHINSPNKANDLGIGMVHQHFMVLDFTTVAENIILGKEPKKFRIDRKRAGQEVQD 123  
 35 Query: 134 LSEKYGLSVNPNAKISDISVGAQORVEILKTLRGADILIFDEPTAVLTPSEIKELMTIM 193  
 +S++YGL ++P AK +DISVG QOR EILKTLRGADILIFDEPTAVLTP EIKELM IM  
 Sbjct: 124 ISDRYGLQIHPEAKADISVGMQQRAEILKTLRGADILIFDEPTAVLTPHEIKELMQIM 183  
 40 Query: 194 KSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGRSVSF 253  
 K+LVKEGKSIILITHKL EI + D+VTVIR+GK I+T+ V + +LA +MVGR VSF  
 Sbjct: 184 KNLVKEGKSIILITHKLKEIMEICDRVTVIRKKGKIKTLDVRDNTQDELASLMVGREVSF 243  
 45 Query: 254 RTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQA 313  
 +TEK+ A P +L++ + V++ R G+ V++LSL V+AGEIVGIAG+DGNGQSELI+A  
 Sbjct: 244 KTEKRAAQPGEVLAIDGITVKDTR-GIETVRDLSSLVKAGEIVGIAGVDGNGQSELIQA 302  
 50 Query: 314 ITGLRKVTSGQIVIKGKDVTKFSSRQITELSVGHVPEDRHRDGLVLDMTMAENLALQTTY 373  
 +TGLRK SG I + GK + + R+ITE +GH+P+DRH+ GLVLD + EN+ LQ+YY  
 Sbjct: 303 VTGLRKTDSGTITLNGKQIQNLTPRKITESGIGHIPQDRHKHGLVLDFFIGENILQSY 362  
 55 Query: 374 KEPLSHKGILNFAKIKEYARQLMTEFVDRGAGEHVLARGFSGGNQQKAI IAREVDRDPDL 433  
 K+P S G+L+ ++ + AR L+TE+DVR E+ AR SGGNQQKAI I RE+DR+PDL  
 Sbjct: 363 KKPYSALGVLHKGEMYKARS LITEYDVRTPEYTHARALSGGNQQKAI IGREIDRNPDL 422  
 Query: 434 LIVSQPTRGLDVGAIEFYIHKRLIEERDKGKAVLVVSFELDEILNLSDRIAVIHDGKIQGI 493  
 LI +QPTRGLDVGAIE++HK+LIE+RD GKAVL++SFEL+EI+NLSDRIAVI +G+I  
 Sbjct: 423 LIAAQPTRGLDVGAIEFVHKRLIEQRDAGKAVLLLSFELEETMNLSDRIAVIFEGRIIAS 482  
 Query: 494 VKPDQTNKQELGILMAG 510  
 V P +T +QELG+LMAG

-1685-

Sbjct: 483 VNPQETTEQELGLLMAG 499

Identities = 75/242 (30%), Positives = 128/242 (51%), Gaps = 24/242 (9%)

Query: 280 GVLAVKNLSLDVRAGEIVGIAGIDGNGQSELIQAITGLRKVTSGQIVIKGKDVTKFSSRQ 339

G++A N++L V+ GEI + G +G G+S L+ + GL + G+I ++G+ V S +

Sbjct: 16 GIVANDNINLQVKKGEIHALLGENGAGKSTLMNVLFGLYQPERGEIRVRGEKVHINSFNK 75

Query: 340 ITELSVGHVPEDRHRDGLVLD-MTMAENLALQTYIKEPLSHKGILNFAKI--KEYARQLM 396

+L +G V H+ +++D T+AEN+ L KEP F +I K +++

Sbjct: 76 ANDLGIGMV----HQHFMLVDTFTVAENIILG---KEPKK-----FGRIDRKRAQGEVQ 122

Query: 397 TEFDVRGAGEHVLARG--FSGGNQOKAIAREVDRDPDLLIVSQPTRGL--DVGAI EYI 451

D G H A+ S G QQ+A I + + R D+LI +PT L ++ + I

Sbjct: 123 DISDRYGLQIHPEAKAADISVGMQORAEILKTLRGADILIFDEPTAVLTPHEIKELMQI 182

Query: 452 HKRLIEERDKGKAVLVVSFELDEILNLSDRIVIHDKGIQGIKVPDQTNKQELGILMAGG 511

K L++E GK+++++ +L EI+ + DR+ VI GK + TN+ EL LM G

Sbjct: 183 MKNLVKE--GKSIILITHKLKEIMEICDRVTVIRKKGKIKTLDVDRDINQDELASLMVGR 239

Query: 512 KI 513

++

Sbjct: 240 EV 241

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4677> which encodes the amino acid sequence <SEQ ID 4678>. Analysis of this protein sequence reveals the following:

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3558(Affirmative) &lt; succ&gt;

bacterial membrane --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 431/511 (84%), Positives = 467/511 (91%), Gaps = 1/511 (0%)

Query: 10 MTQNVIEKKEITKKFGDFVANDHINLTVEKGEIHALLGENGAGKSTLMNMLAGLLEPTDG 69

MTQ+VIEK+EITKKFGDFVANDHINL V KGEIHALLGENGAGKSTLMNMLAGLLEPT G

Sbjct: 7 MTQHVIEMREITKKFGDFVANDHINLVNRKGEIHALLGENGAGKSTLMNMLAGLLEPTSG 66

Query: 70 QIFINGQPVTIDSPSKSSQLGIGMVHQHFMLVEAFTVAENIVLGNETTQNGVLDIKTAAK 129

+I IN +PV IDSPSKS++LGIGMVHQHFMLVEAFTVAENI+LGNE +NG LD+ A+K

Sbjct: 67 EIVINDKVPQIDSPSKSAKLGIGMVHQHFMLVEAFTVAENIILGNEVVKNGLDLNQASK 126

Query: 130 EIKELSEKYGLSVNPNAKISDISVGAQQRVEILKTLRGADILIFDEPTAVLTPSEIKEL 189

+IK LSEKYGL++NP+AK+SDISVGAQQRVEILKTLRGADILIFDEPTAVLTP+EIKEL

Sbjct: 127 DIKVLSEKYGLAINPSAKVSDISVGAQQRVEILKTLRGADILIFDEPTAVLTPAEIKEL 186

Query: 190 MTIMKSLVKEGKSIILITHKLDEIRAVADKVTVIRRGKSIETVPVAGASSQQLAEMMVGR 249

MTIMK+LVKEGKSIILITHKLDEIRAVAD+VTVIRRGKSIETV VAGA+SQ LAEMMVGR

Sbjct: 187 MTIMKNLVKEGKSIILITHKLDEIRAVADRVTVIRRGKSIETVDVAGATSQDLAEMMVGR 246

Query: 250 SVSFRTEKKEANPTDIILSVKDLVVEENRGGVLAVKNLSLDVRAGEIVGIAGIDGNGQSE 309

SVSF T KK A P D++LS+K+L V+ENR GV AVK LSLDVRAGEIVGIAGIDGNGQSE

Sbjct: 247 SVSFTTSKKAEPKDVLSIKNLEVDENR-GVPAVKGLSLDVRAGEIVGIAGIDGNGQSE 305

Query: 310 LIQAITGLRKVTSGQIVIKGKDVTKFSSRQITELSVGHVPEDRHRDGLVLDMTMAENLAL 369

LIQAITGLRKV SG I+IK +VT SSR+ITELSVGHVPEDRHRDGL+LD+++AEN AL

Sbjct: 306 LIQAITGLRKVSGSIMIKNNEVTHLSSRKITELSVGHVPEDRHRDGLILDLSLAENTAL 365

Query: 370 QTYIKEPLSHKGILNFAKIKEARQLMTEFDVRGAGEHVLARGFSGGNQOKAIAREVDR 429

QTYIK+PLS GILN+ KI +YARQLM EFDVRGA E V ARGFSGGNQOKAIAREVDR

Sbjct: 366 QTYIKPLSQNGIILNYTKINDYARQLMKEFDVRGANELVPARGFSGGNQOKAIAREVDR 425

Query: 430 DPDLLIVSQPTRGLDVGAI EYIHKRLIEERDKGKAVLVVSFELDEILNLSDRIVIHDKG 489

-1686-

DPDLLIVSQPTRGLDVGAI EYIHKRLI+ERDKGKAVLVVSFELDEILNLSDR IAVIHDGK  
 Sbjct: 426 DPDLLIVSQPTRGLDVGAI EYIHKRLIKERDKGKAVLVVSFELDEILNLSDR IAVIHDGK 485

Query: 490 IQGIVKPDQTNKQELGILMAGGKIEKEERDV 520  
 IQGIV P+ TNKQELGILMAGG I KEE V  
 Sbjct: 486 IQGIVSPENTNKQELGILMAGGSIHKKEGHV 516

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 1521

A DNA sequence (GBSx1612) was identified in *S.agalactiae* <SEQ ID 4679> which encodes the amino acid sequence <SEQ ID 4680>. Analysis of this protein sequence reveals the following:

Possible site: 22  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)  
 [Bacillus subtilis]  
 Identities = 164/335 (48%), Positives = 224/335 (65%), Gaps = 10/335 (2%)

Query: 18 LAACGHRGASKSGGKS-DSLKVAMVTD TGGVDDKSFNQSGWEGMQAWGKKNGLKKGA-GF 75  
 L ACG+ S G+ + VAMVTD GGVDDKSFNQSG WEG+QA+GK+NGLKKG G+  
 Sbjct: 11 LGACGNSEKSSSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKENGKKGKNGY 70

Query: 76 DYFQSASESDYATNLD TAVSSGYKLIFGIGFSLHDAIDKAADNNKDVNYVIVDDVIKGD 135  
 DY QS S++DY TNL+ + LI+G+G+ + D+I + AD K+ N+ I+D V+ KD  
 Sbjct: 71 DYLSKSDADYTTNLNKLARENFDLIYGVGYLMEDSISEIADQRKNTNF AII DAVVD-KD 129

Query: 136 NVASVVFADNESAYLAGIAAAKTTKT VTFVGGMESEVITRFEKGFEAGVKSVDKSIKI 195  
 NVAS+ F + E ++L G+AAA ++K+ +GFVGGMESE+I +FE GF AGV++V+ +  
 Sbjct: 130 NVASITFKEQEGSFLVGVAAALSSKSGKIGFVGGMESELIKKFEVGFAGVQAVNPKAVV 189

Query: 196 KVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSEAKSRNESLKEADKVWVL 255  
 +V YAG F A GK A + Y SG D++Y AG TG GVF+EAK+ + + D VWV+  
 Sbjct: 190 EVKYAGGFDKADV GKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKEDPKRD-VWVI 248

Query: 256 GVD RDQA AEGKYTSKDGKASNFVLASSI KEV GKSVELIATKTSKGKFPGGNVT TYGLKD G 315  
 GVD+DQ AEG+ +G N L S +K+V VE + K S GKFP GG TYGL  
 Sbjct: 249 GVDKDQY AEGQV---EGTDDNVTLTSMVKKVDTVVEDVTKKASDGKFPGETLTYGLDQD 305

Query: 316 GV DIATT--NLSDDAVKAIKEAKAIISGDIKVPS 348  
 GV I+ + NLSDD +KA+ + K KII G +++P+  
 Sbjct: 306 GVGISPSKQNLSDDV IKAVDKWKKKIIDG-LEIPA 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 861> which encodes the amino acid sequence <SEQ ID 862>. Analysis of this protein sequence reveals the following:

possible site: 19  
 >>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1687-

Identities = 275/351 (78%), Positives = 312/351 (88%), Gaps = 3/351 (0%)

Query: 1 MNKKIAGIGLASIAVLSLAACGHRGASKSG--GKSDSLKVAMVTDITGGVDDKSFNQSGWE 58  
 MNKK G+GLAS+AVLSLAACG+RGASK G GK+D LKVAMVTDITGGVDDKSFNQS WE  
 Sbjct: 1 MNKKFIGLGLASVAVLSLAACGNRGASKGGASGKTD-LKVAMVTDITGGVDDKSFNQSAWE 59

Query: 59 GMQAWGKKNGLLKKGAGFDYFQSASESDYATNLDITAVSSGYKLIFGIGFSLHDAIDKAADN 118  
 G+Q+WGK+ GL+KG GFDYFQS SES+YATNLDITAVS GY+LI+GIGF+L DAI KAA +  
 Sbjct: 60 GLQSWGKEMGLQKGTGFDYFQSTSESEYATNLDITAVSSGYQLIYIGIGFALKDAIAKAAGD 119

Query: 119 NKDVNYVIVDDVIKGDNVASVVFADNESAYLAGIAAAKTTTKTKTVGFVGGMESEVITRF 178  
 N+ V +VI+DD+I+GKDNVASV FAD+E+AYLAGIAAAKTTTKTKTVGFVGGME VITRF  
 Sbjct: 120 NEGKVFVIIDDIIEGKDNVASVTFADHEAAYLAGIAAAKTTTKTKTVGFVGGMEGTVITRF 179

Query: 179 EKGFEAGVKSVDKSIKIKVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSE 238  
 EKGFEAGVKSVD +I++KVDYAGSFGDAAKGKTIAAAQYA+GAD++YQ AGGTGAGVF+E  
 Sbjct: 180 EKGFEAGVKSVDVTIQKVDYAGSFGDAAKGKTIAAAQYAAGADVIVYQAGGTGAGVFNE 239

Query: 239 AKSRNESLKEADKVWVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKSVELIATKTS 298  
 AK+ NE EADKVWV+GVDRDQ EGKYTSKDGK +NFVLASSIKEVGK+V+LI + +  
 Sbjct: 240 AKAINERSEADKVWVIGVDRDQKDEGKYTSKDGKEANFVLASSIKEVGKAVQLINKQVA 299

Query: 299 KGKFPGGNVTTYGLKDGVDIATTNLSDDAVKAIKEAKAKIISGDIKVP SK 349  
 KFPGG T YGLKDGGV+IATTN+S +AVKAIKEAKAKI SGDIKVP K  
 Sbjct: 300 DKKFPGGKTTVYGLKDGGVETIATTNVSKEAVKAIKEAKAKIKSGDIKVP EK 350

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 9061> which encodes amino acid sequence <SEQ ID 9062>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> May be a lipoprotein

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

Score = 414 bits (1052), Expect = e-117  
 Identities = 196/347 (56%), Positives = 253/347 (72%), Gaps = 2/347 (0%)

Query: 1 MNKKVMSLGLVSTALFTLGGCTNNSAKQT--TDNSLKIAMITNQTGIDDKSFNQSAWEG 58  
 MNKK+ +GL S A+ +L C + A ++ +SLK+AM+T+ G+DDKSFNQS WEG+  
 Sbjct: 1 MNKKIAGIGLASIAVLSLAACGHRGASKSGGKSDSLKVAMVTDITGGVDDKSFNQSGWEGM 60

Query: 59 QAWGKENKLEKGGYDYFQSANESFTTNLESATVNGYNLVFGIGFPLHDAVEKVAANNP 118  
 QAWGK+N L+KG G+DYFQSA+ES++ TNL++AV++GY L+FGIGF LHDA++K A NN  
 Sbjct: 61 QAWGKKNGLLKKGAGFDYFQSASESDYATNLDITAVSSGYKLIFGIGFSLHDAIDKAADNNK 120

Query: 119 DNHFAIVDDVIKQKNVASITFSDHEAAYLAGVXXXXXXXXXXQVGFVGGMEGDVVKRFEK 178  
 D ++ IVDDVIK+ NVAS+ F+D+E+AYLAG+ VGFVGGME +V+ RFEK  
 Sbjct: 121 DVNYVIVDDVIKGDNVASVVFADNESAYLAGIAAAKTTTKTKTVGFVGGMESEVITRFEK 180

Query: 179 GFEAGVKSVDVTIKVRVAYAGSFXXXXXXXXXXXXXXXXXEGADVIVYHAAGGTGAGVFSEAK 238  
 GFEAGVKSVD +IK++V YAGSF GAD++Y AGGTGAGVFSEAK  
 Sbjct: 181 GFEAGVKSVDKSIKIKVDYAGSFGDAAKGKTIAAAQYASGADIVYQVAGGTGAGVFSEAK 240

Query: 239 SINEKRKEEDKVWVIGVDRDQSEDGKYTTKDGKASNFVLTSIIEVGKALVKVAVKTS 298  
 S NE KE DKVWV+GVDRDQ+ +GKYT+KDGK++NFVL SSIIEVGK++ +A KTS+  
 Sbjct: 241 SRNESLKEADKVWVLGVDRDQAAEGKYTSKDGKASNFVLASSIKEVGKSVELIATKTSKG 300

Query: 299 QFPGGQITTFGLKEGGVSLTTFDALTDQDTXXXXXXXXXXXXXGTTTVP 345  
 +FPGG +TT+GLK+GGV + T L+ D G I VP  
 Sbjct: 301 KFPGGNVTTYGLKDGVDIATTNLSDDAVKAIKEAKAKIISGDIKVP 347

-1688-

SEQ ID 4680 (GBS211) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 6; MW 40kDa).

The GBS211-His fusion product was purified (Figure 205, lane 8) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 259A) and FACS (Figure 259B). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1522

A DNA sequence (GBSx1613) was identified in *S.agalactiae* <SEQ ID 4681> which encodes the amino acid sequence <SEQ ID 4682>. This protein is predicted to be cytidine deaminase (cdd). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2112(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9801> which encodes amino acid sequence <SEQ ID 9802> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB51906 GB:AJ237978 cytidine deaminase [Bacillus psychrophilus]
Identities = 66/114 (57%), Positives = 81/114 (70%)

Query: 26 KASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIFKAVSEGYKDFS 85
      KA E AYVPYSKFPVGAAL +G I+ GCN+EN +Y + NCAERTA FKAVS+G + F
Sbjct: 12 KAREQAYVPYSKFPVGAALLAEDGTIYHGCNIENSAYSMTNCAERTAFFKAVSDGVRSEFK 71

Query: 86 EIAIYGNTEEPISPCGACRQVMVEFFNKNKAVTLIAKNGKTVETTVEGELLPSYF 139
      +A+ +TE P+SPCGACRQV+ EF N + V L G ETTV +LLP +F
Sbjct: 72 ALAVVADTEGPVSPCGACRQVIAEFCNGSMPVYLTLNKGDIETTVAKLLPGAF 125
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4683> which encodes the amino acid sequence <SEQ ID 4684>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0041(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15143 GB:Z99120 similar to ABC transporter (lipoprotein)
[Bacillus subtilis]
Identities = 152/339 (44%), Positives = 223/339 (64%), Gaps = 11/339 (3%)

Query: 8 LGLVSTALFTLGGCTNN---SAKQTTDNSLKIAMITNQTGIDDKSFNQSAWEGLOAWGKE 64
      + LV A LG C N+ S N +AM+T+ G+DDKSFNQSAWEG+QA+GKE
Sbjct: 1 MSLVIAAGTILGACGNSEKSSGSGEGKNKFSVAMVTDVGGVDDKSFNQSAWEGIQAFGKE 60
```

-1689-

Query: 65 NKLEKGGK-GYDYFQSANESFTTNLESAVINGYNLVFGIGFPLHDAVEKVAANNPDNHFA 123  
 N L+KGK GYDY QS +++++TINL ++L++G+G+ + D++ ++A + +FA  
 Sbjct: 61 NGLKKGKNGYDYLQSKSDADYTINLNLKLAENFDLIYGVGYLMEDSISEIADQRKNTNFA 120

5 Query: 124 IVDDVIKQKQNVASITFSDEHAYLAGVAAAKTTKTQVGFVGGMEGDVVKRFEKGFAG 183  
 I+D V+ + NVASITF + E ++L GVAAA ++K+ ++GFVGGME +++K+FE GF AG  
 Sbjct: 121 IIDAVVD-KDNVASITFKEQEGSFLVGVAALSSKSGKIGFVGGMESELIKKEFEVGFAG 179

10 Query: 184 VKSVDDTIKVRVAYAGSFADAAGKGTIAAAQYAEADVIYHAAGGTGAGVFSEAKSINEK 243  
 V++V+ V V YAG F A GK A + Y G DVIYH+AG TG GVF+EAK++ ++  
 Sbjct: 180 VQAVNPKAVVEVKYAGGFDKADVGKATAESMYKSGVDVIYHSAGATGTGVFTEAKNLKKE 239

15 Query: 244 RKEEDKVVWIGVDRDQSEDGKYTTKDGKSAFVLTSSIKEVGKALVKVAVKTSQDQFPGG 303  
 + D VVWIGVD+DQ +G+ +G N LTS +K+V + V K S+ +FPFGG  
 Sbjct: 240 DPKRD-VVWIGVDKDQYAEQV---EGTDDNVLTSMVKKVDTVVEDVTKKASDGKFPFGG 295

20 Query: 304 QITTFGLKEGGVSLTTDA--LTQDTKKAIEAAKKAIEG 340  
 + T+GL + GV ++ L+ D KA++ KK II+G  
 Sbjct: 296 ETLYGLDQDGVGISPSKQNLSDDDVIKAVDKWKKIIDG 334

An alignment of the GAS and GBS proteins is shown below.

Identities = 88/128 (68%), Positives = 107/128 (82%)

25 Query: 15 MGNIELKKLAVKASENAYVPYSKFPVGAALRTAEGKIFTGCNVENISYGLANCAERTAIF 74  
 MG +L AV+ASE AYVPYS FVGAAL+T +G I+TCN+EN+S+GL NC ERTAIF  
 Sbjct: 1 MGTDLVSCAVQASEYAYVPYSHFPVGAALKTKDGTIYTCNIENVSFGLTNCGERTAIF 60

30 Query: 75 KAVSEGYKDFSEIAIYGNTEEPISPCGACRQVMVEFFNKNKAVTLIAKNGKTVEITVDEL 134  
 KA+S+G+K+ EIAIYG T +P+SPCGACRQVM EFF+ ++ VTIAKNG+TVE TVG+L  
 Sbjct: 61 KAISDGHKELVEIAIYGETMQPVSPCGACRQVMAEFFDPSSLVTLIAKNGQTVEMTVGDL 120

35 Query: 135 LPYSFVDL 142  
 L YSF DL  
 Sbjct: 121 LLYSFTDL 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1523

40 A DNA sequence (GBSx1614) was identified in *S. agalactiae* <SEQ ID 4685> which encodes the amino acid sequence <SEQ ID 4686>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2979(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 A related GBS nucleic acid sequence <SEQ ID 9799> which encodes amino acid sequence <SEQ ID 9800> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:CAB11882 GB:Z99104 alternate gene name: ybaA~similar to  
 hypothetical proteins [Bacillus subtilis]  
 Identities = 90/201 (44%), Positives = 144/201 (70%), Gaps = 5/201 (2%)

Query: 1 MANMYTYTENPNVEHDIHENVKLLGESFSFLTDAGVFSKRMIDYGSQVLNLSLHF-EKNK 59  
 M+ YY+E P+V+ + +L + F+F +D+GVFSK+ +D+GS++L++S E  
 Sbjct: 1 MSEHYSEKPSVKSNNKQTSFRLRNKDFTFSTDSGVFSKKEVDGSRLLIDSFEEPEVEG 60

-1690-

Query: 60 SLLDLGCGYGPLGISLAK-VQGVKATMVDINTRALELAKKNATRNGVV-VEVFQSNIIYEN 117  
 +LD+GCGYGP+G+SLA + M+D+N RA+EL+ +NA +NG+ V+++QS+++ N  
 Sbjct: 61 GILDVGCYGPGLGLSLASDFKDRITIHMDVNERAVELSNENAEQNGITNVKIYQSDLFEN 120

Query: 118 I--SKTFDYIISNPPIRAGKQVHSHIIEESICYLNTGGSLTIVIQKKQGAPSAKAKMLDT 175  
 + ++TF I++NPPIRAGK+VVH+I E+S +L G L IVIQKKQGAPSA K+ +  
 Sbjct: 121 VDSAQTFASILTNPPIRAGKKVHAI FEKSAEHLKASGELWIVIQKKQGAPSAIEKLEEL 180

Query: 176 FGNC DILKKDKGYI LRSEKV 196  
 F +++K KGYI++++KV  
 Sbjct: 181 FDEVSVVQKKKGYI IAKKV 201

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4687> which encodes the amino acid sequence <SEQ ID 4688>. Analysis of this protein sequence reveals the following:

Possible site: 49  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4232(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 139/195 (71%), Positives = 165/195 (84%)  
 Query: 1 MANMYITENPNVEHDIHNLVKKLLGESFSFLTDAGVFSKRMIDYGSQVLLNSLHFEKNKS 60  
 M MYE ENP+ HDIHE+ V+LL F+FLTD+GVFSK+M+D+GSQVLL +L+F++N+  
 Sbjct: 12 MTKMYIDENPDSDLHDIHEVKVELLNHPFTFLTDAGVFSKRMVDFGSQVLLKTLNFKENER 71  
 Query: 61 LLDLGCGYGPLGISLAKVQGVKATMVDINTRALELAKKNATRNGVVVEVFQSNIIYENISK 120  
 +LDLGCGYGPLGISLAKVQ V AT+VDIN RAL+LA+KNAT N V V +FQSNIIYENIS  
 Sbjct: 72 VLDLGCGYGPLGISLAKVQRVDATLVDINNRLDLARKNATNNQVAVTIFQSNIIYENISG 131  
 Query: 121 TFDYIISNPPIRAGKQVHSHIIEESICYLNTGGSLTIVIQKKQGAPSAKAKMLDTFGNCD 180  
 F++IISNPPIRAGK+VVHSHIIE+SI +L G LTIVIQKKQGAPSAKAKM FGN +  
 Sbjct: 132 HFEHIISNPPIRAGKRVHSHIIEKSIDFLVVGDLTIVIQKKQGAPSAKAKMATIFGNVE 191  
 Query: 181 ILKKDKGYI LRSEK 195  
 IL+KDKGYI+LRS K  
 Sbjct: 192 ILRKDKGYI VLRSEK 206

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1524

A DNA sequence (GBSx1615) was identified in *S.agalactiae* <SEQ ID 4689> which encodes the amino acid sequence <SEQ ID 4690>. This protein is predicted to be pantothenate kinase (coaA). Analysis of this protein sequence reveals the following:

Possible site: 28  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.5021(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06594 GB:AP001516 pantothenate kinase [Bacillus halodurans]  
 Identities = 140/307 (45%), Positives = 203/307 (65%), Gaps = 5/307 (1%)

-1691-

Query: 4 EFINFDRISRENWKDLHQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQIYKR 63  
 +F + +SR WK L + S + E+ELE + LN+ I + +V DIY+PL L+ ++  
 Sbjct: 8 DFFPYTVLSRSQWKSRLKASSLPINEQELEQLVGLNEPITLNEVADIYVPLAELLHVHAT 67

Query: 64 SQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELVTTD 123  
 + + L K F + PFIIG++GSVAVGKSTT+RLLQ L+ + HV+LVTTD  
 Sbjct: 68 AYQRLQQQKRGFHFGKNSPFIIGLAGSVAVGKSTTARLLQKLLKAWPEHHHVDLVTTD 127

Query: 124 GFLYPNEKLIQNGILNRKGFPEESYDMESLLNFDLTIKNGIT-AKIPYSHEIYDIVPNQL 182  
 GFLYPNE L G++++KGFPESYD+ +L+ FL +K G K P+YSH Y+IV  
 Sbjct: 128 GFLYPNETLEARGLMDKKGFPEESYDLPALIRFLSDVKAGEPYVKAPVYSHLTYNIVEGDY 187

Query: 183 QTIETPDFLILEGINVFQ-NQQNHRL---YMNDYFDFSIIYIDAENKQIEEWYLQRFNSLL 238  
 Q + PD +I+EGINV Q N++NH + +++D+FDFSIIY+DA+ +QI +WY++RF L  
 Sbjct: 188 QVVHEPDIVIVEGINVLQVNRNHHIPNVFVSDFFDFSIIYVDAKEEQILQWYIERFKLLQ 247

Query: 239 QLAEADPSNYHFKFTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIIHKGKHHK 298  
 A DP++Y+H+F + +A + A IWK IN VNL + I PT++RAD ++ KG HH  
 Sbjct: 248 NTAQDPNSYFHRFRHLSEVEAEQFATSIWKNINGVNLHENILPTKHRADLVLQKGPHEF 307

Query: 299 IDEIYLK 305  
 IDE+ L+  
 Sbjct: 308 IDEVKLR 314

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4691> which encodes the amino acid sequence <SEQ ID 4692>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4790(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 219/306 (71%), Positives = 269/306 (87%)

Query: 1 MNNEFINFDRISRENWKDLHQSQALLTEKELESIKSLNDNINIQDVIDIYLPLINLIQI 60  
 M+NEFINF++ISRE+WK LHQ+++ALLT++EL+SI SLNDNI+I DVIDIYLPLINLIQ+  
 Sbjct: 1 MSNEFINFEKISRESWKTLHQKAKALLTQEELKSITSLNDNISINDVIDIYLPLINLIQV 60

Query: 61 YKRSQENLSFSKAIFLKKENYQRPFIIGISGSVAVGKSTTSRLQLLISRTFKDSHVELV 120  
 YK +QENLSFSK++FLKK+ RPFIIIGISGSVAVGKSTTSRLQLL+SRT +S VELV  
 Sbjct: 61 YKIAQENLSFSKSLFLKKDIQLRPFIIIGISGSVAVGKSTTSRLQLLRSRTHPNSQVELV 120

Query: 121 TTDGFLYPNEKLIQNGILNRKGFPEESYDMESLLNFDLTIKNGITAKIPYSHEIYDIVPN 180  
 TTDGFLYPN+ LI+ G+LNRKGFPEESY+ME LL+FLD+IKNG TA P+YSH+IYDI+PN  
 Sbjct: 121 TTDGFLYPNQFLIEQGLLNRKGFPEESYNMELLLDFLDSIKNGQTAFAPVYSHDIYDIIPN 180

Query: 181 QLQTIETPDFLILEGINVFQNNQNHRLYMNDYFDFSIIYIDAENKQIEEWYLQRFNSLLIQL 240  
 Q Q+ PDFLI+EGINVFNQNN+RLYM+DYFDFSIIYIDA++ IE WY++RF S+L+L  
 Sbjct: 181 QKQSFNNPDFLIVEGINVFQNNQNRLYMSDYFDFSIIYIDADSSHIETWYIERFLSILKL 240

Query: 241 AEADPSNYHFKFTQIPPHKAMELAKDIWKTINLVNLEKYIEPTRNRADFIIHKGKHHKID 300  
 A+ DP NYY ++ Q+P +A+ A+++WKT+NL NLEK+IEPTRNRA+ I+HK HKID  
 Sbjct: 241 AKRDPHNYAQAQLPRSEATAFARNVWKTIVNLENLEKFIETPTRNRAELILHKSADHKID 300

Query: 301 EIYLKK 306  
 EIYLKK  
 Sbjct: 301 EIYLKK 306

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1692-

**Example 1525**

A DNA sequence (GBSx1616) was identified in *S.agalactiae* <SEQ ID 4693> which encodes the amino acid sequence <SEQ ID 4694>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3866(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05058 GB:AP001511 ribosomal protein S20 (BS20) [Bacillus halodurans]  
Identities = 47/86 (54%), Positives = 59/86 (67%), Gaps = 7/86 (8%)

Query: 3 VKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEA---NPSEELYRA----ASSS 55  
+K ANIKSAIKR + N K+ +N++ KSA+RTAIK FEA N E +A A+  
Sbjct: 1 MKGNANIKSAIKRVKTNEKRRIQNASVKSALRTAIKQFEAKVENNDAAKAAAFVEATKK 60

Query: 56 IDKAASKGLIHTNKASRDKARLATKL 81  
+DKAA+KGLIH N ASR K+RLA KL  
Sbjct: 61 LDKAANKGLIHKNAASRQKSRLAKKL 86

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4695> which encodes the amino acid sequence <SEQ ID 4696>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3872(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 76/82 (92%), Positives = 78/82 (94%)

Query: 1 MEVKTLANIKSAIKRAELNVKQNEKNSAQKSAMRTAIKAFEANPSEELYRAASSSIDKAA 60  
+EVKTLANIKSAIKRAELNVK NEKNSAQKSAMRTAIKAFEANPSEEL+RAASSSIDKA  
Sbjct: 1 LEVKTLANIKSAIKRAELNVKANEKNSAQKSAMRTAIKAFEANPSEELFRAASSSIDKAE 60

Query: 61 SKGLIHTNKASRDKARLATKLG 82  
SKGLIH NKASRDKARLA KLG  
Sbjct: 61 SKGLIHKNKASRDKARLAALKG 82

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1526**

A DNA sequence (GBSx1617) was identified in *S.agalactiae* <SEQ ID 4697> which encodes the amino acid sequence <SEQ ID 4698>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -10.99 Transmembrane 31 - 47 ( 25 - 51)

----- Final Results -----

bacterial membrane --- Certainty=0.5394(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1693-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC35851 GB:AF086736 amino acid-binding protein Abp  
[Streptococcus uberis]  
Identities = 169/269 (62%), Positives = 203/269 (74%), Gaps = 2/269 (0%)

10 Query: 29 KNILLTIIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGD 88  
K ILLT + + L ACG S+ A D W+ Y+KEK IT+GFDNTFVPMGF+ SG  
Sbjct: 4 KKILLTTLALASTLFLVACGKSSA--AKTDQWDTYKKEKSITLGFNTFVPMGFKDES GK 61

15 Query: 89 YTGFDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTN 148  
TGFD++LA AVF+EYGI VK+QPINWD+KETEL NG ID+IWNNGYS T ER KVAFA+  
Sbjct: 62 NTGFDVELAKAVFQYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFST 121

Query: 149 PYMNNHQVIVTKTSSHINSIKDMKGKGLGAQSGSSGFDFAFNAKPDILKKFVKGEAVQYD 208  
PYM N QV+VTK SS+I S MKGK LGAQSGSSG+DAF + P +LK VK +A QY+  
Sbjct: 122 PYMKNEQVLVTKSSNITSFAAMKGVGLGAQSGSSGYDAFTSNPKVLKDIVKDNDAQY 181

20 Query: 209 TTFQALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAQGENFVVGARKVDRRL 268  
TF QA IDLKN+RIDGLLID+VYANYYLKQEG + Y VK+ + GE+F VG RK D+ L  
Sbjct: 182 TTFQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDKIL 241

25 Query: 269 IEKINKAFKQLHNKGRFQKISYKWFGEDEV 297  
++ IN AF +L+ G+FQ+IS KWFGEDEV  
Sbjct: 242 LKNINSAFTKLYKTGKFQEISQKWFGEDEV 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4699> which encodes the amino acid sequence <SEQ ID 4700>. Analysis of this protein sequence reveals the following:

30 Possible site: 21

>>> May be a lipoprotein

35 ----- Final Results -----  
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 >GP:AAC35851 GB:AF086736 amino acid-binding protein Abp  
[Streptococcus uberis]  
Identities = 176/277 (63%), Positives = 220/277 (78%), Gaps = 1/277 (0%)

45 Query: 1 MIIKRTVAILAIASSFFLVACQATKSLKSGDANGVYQKQKSITVGFNTFVPMGYKDES 60  
M +KK + LA+AS+ FLVAC + + K+ D W Y+K+KSIT+GFDNTFVPMG+KDES  
Sbjct: 1 MNLKKILLTTLALASTLFLVACGKSSAAKT-DQWDTYKKEKSITLGFNTFVPMGFKDES 59

50 Query: 61 GRCKGFDIDLAKVVFHQYGLKVNFOAINWDMKEAELNNGKIDVIWNGYSITKERQDKVAF 120  
G+ GFD++LAK VF +YG+KV FQ INWD+KE EL NGKID+IWNNGYS+TKERQ KVAFA  
Sbjct: 60 GKNTGFDVELAKAVFQYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAF 119

55 Query: 121 TDSYMRNEQIIVVKKRSDIKTISDMKHVKVLGAQSASSGYDSLRLTPKLLKDFIKNKDANQ 180  
+ YM+NEQ++V KK S+I + + MK KVLGAQS SSGYD+ PK+LKD +K+ DA Q  
Sbjct: 120 STPYMKNEQVLVTKSSNITSFAAMKGVGLGAQSGSSGYDAFTSNPKVLKDIVKDNDAQ 179

Query: 181 YETFTQAFIDLKSDRIDGILIDKVYANYYLAKQGLENYRMIPTTFENEAFSVGLRKEDK 240  
YETF QAFIDLK+DRIDG+LIDKVYANYYL +EG+L NY ++ + F+ E F+VG+RKEDK  
Sbjct: 180 YETFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVGVRKEDK 239

60 Query: 241 TLQAKINRAFRVLYQNGKFQAISEKWFGEDEVATANIK 277  
L IN AF LY+ GKFO IS+KWFGEDEVAT N+K  
Sbjct: 240 ILLKNINSAFTKLYKTGKFQEISQKWFGEDEVATENVK 276

-1694-

An alignment of the GAS and GBS proteins is shown below.

Identities = 151/266 (56%), Positives = 189/266 (70%), Gaps = 4/266 (1%)

```

5  Query: 32  LLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTG 91
      +L I   F++  AC + K +   D W  YQK+K IT+GFDNTFVPMG++ SG  G
Sbjct: 10  ILAIASSFFLV---AC-QATKSLKSGDAWGVYQKQKSITVGFDTNFVPMGYKDESGRCKG 65

Query: 92  FDIDLANAVFKEYGISVKWQPINWDMKETELNNGNIDLIWNGYSKTAERAKKVAFTNPYM 151
      FDIDLA VF +YG+ V +Q INWDMKE ELNNG ID+IWNGYS T ER  KVAFT+ YM
10 Sbjct: 66  FDIDLAKEVFHQYGLKVNFAQINWDMKEAELNNGKIDVIWNGYSITKERQDKVAFTDSYM 125

Query: 152  NNHQVIVTKTSSHINSIKDMKGKKLGAQSGSSGFDAFNAKPDILKKFVGKKEAVQYDTFT 211
      N Q+IV K  S I +I DMK K LGAQS SSG+D+   P +LK F+K K+A QY+TFT
15 Sbjct: 126  RNEQIIVVKRSDIKTISDMKHKVLGAQSASSGYDSLLRTPKLLKDFIKNKDANQYETFT 185

Query: 212  QALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKTAQGENFVVGARKVDRRLIEK 271
      QA IDLK++RIDG+LID+VYANYYL +EG ++ Y  + T ++ E F VG RK D+ L  K
20 Sbjct: 186  QAFIDLKSDRIDGILIDKVYANYYLAKEGQLENYRMIPTTFENEAFSVGLRKEDKTLQAK 245

Query: 272  INKAFKQLHNKGRFQKISYKWFGEVDV 297
      IN+AF+ L+  G+FQ IS KWFG+DV
25 Sbjct: 246  INRAFRVLYQNGKFQAISEKWFGDDV 271

```

25 A related GBS gene <SEQ ID 8833> and protein <SEQ ID 8834> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: 22  Crend: 4
Sequence Pattern: CGMS
SRCFLG: 0
30 McG: Length of UR: 22
      Peak Value of UR: 3.05
      Net Charge of CR: 2
McG: Discrim Score: 11.16
GvH: Signal Score (-7.5): -1.96
      Possible site: 24
35 >>> May be a lipoprotein
Amino Acid Composition: calculated from 23
ALOM program count: 0 value: 8.96 threshold: 0.0
      PERIPHERAL Likelihood = 8.96 68
      modified ALOM score: -2.29
40 *** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50 62.2/75.8% over 270aa
      Streptococcus uberis
      GP|3603430| amino acid-binding protein Abp Insert characterized

ORF00904(385 - 1203 of 1503)
55 GP|3603430|gb|AAC35851.1|AF086736(4 - 274 of 277) amino acid-binding protein Abp
{Streptococcus uberis}
%Match = 34.8
%Identity = 62.1 %Similarity = 75.7
Matches = 169 Mismatches = 65 Conservative Sub.s = 37

60 153      183      213      243      273      303      333      363
      FHYLGGKSNVSH*LVR**LIHRLLVMMSQLALLIQSCVKK*KN*FYKIEKQV*HKL**HMI FNLLKVYLIRFSKLILSRL

393      423      453      483      513      543      573      603
      GGRLLTHKNILLTIIFGLFMIILSACGMSNKEMAGIDNWEHYQKEKKITIGFDNTFVPMGFESRSGDYTGFDIDLANAVF

```

-1695-

```

      : | ||| :      : | ||| :      | | : | ||| ||: || || || || || || : || || || || : || ||
      MNLKKILLTTLALASTLFLVACGKSS--AAKTDQWDITYKEKSITLGFNTFVPMGFKDESGKNTGFDVELAKAVF
      10          20          30          40          50          60          70

5    633      663      693      723      753      783      813      843
      KEYGISVKWQPINWDMKETELNNGNIDLWNGYSKTAERAKKVAFTNPYMMNHQVIVTKTSSHINSIKDMKGKGLGAQSG
      : ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| ||| : ||| |||
      QEYGIKVKFQPINWDLKETELKNGKIDMIWNGYSVTKERQAKVAFSTPYMKNEQVLVTKKSSNITSFAAMKGVLGQAQSG
      90          100         110         120         130         140         150

10   873      903      933      963      993      1023      1053      1083
      SSGFDAFNAKPDILKKFVKGKEAVQYDTFTQALIDLKNNRIDGLLIDEVYANYYLKQEGNIKAYYFVKITAYQGENFVGA
      ||| : ||| : | : || || : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
      SSGYDAFTSNPKVLKDIVKNDATQYETFIQAFIDLKNDRIDGLLIDKVYANYYLKQEGELTNYNIVKSEFDGEDFAVG
      15   170         180         190         200         210         220         230

      1113      1143      1173      1203      1233      1263      1293      1323
      RKVDRRLIEKINKAFKQLHNKGRFQKISYKWFGEVDVYSKE*KTRNFS*SFILRKN*IKNIDISDVF*VN*PSLVARRALS
      || : || : || || : || : || || || || || || || || || || || || || || || || || || || || ||
      RKEDKILLKNINSAFTKLYKTGKFQEISQKWFGEVDVATENVKK
      20   250         260         270

```

SEQ ID 8834 (GBS225) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 10; MW 32kDa). The GBS225-His fusion product was purified (Figure 205, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 266), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1527

A DNA sequence (GBSx1618) was identified in *S.agalactiae* <SEQ ID 4701> which encodes the amino acid sequence <SEQ ID 4702>. This protein is predicted to be arginine ABC transporter, ATP-binding protein (glnQ). Analysis of this protein sequence reveals the following:

Possible site: 39

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

      bacterial cytoplasm --- Certainty=0.3229(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB49429 GB:U73111 glutamine transport ATP-binding protein GLNQ  
[Salmonella typhimurium]

Identities = 94/210 (44%), Positives = 146/210 (68%), Gaps = 3/210 (1%)

```

Query: 1  MLELKNISKCYGQKEIFKDFNLTVEEGKILSLVGPSSGGKTTLLRLAGLEKIDSGTIVH 60
      M+E KN+SK +G ++ + +L + +G+++ ++GPSG GK+TLLR + LE+I SG ++
Sbjct: 1  MIEFKNVSKHFQPTQVLHNIDLNIRQGEVVVVIIGPSGSGKSTLLRCINKLEEITSGDLIV 60

```

```

Query: 61  DGKEVS---VDHLETLNLLGFFVFQDFQLFPHLTVLDNLILSPVKTMGLSKELAKEKALVL 117
      DG +V+ VD G VFQ F LFPHLT L+N++ P++ G+ KE A+++A L
Sbjct: 61  DGLKVNDPKVDERLIRQEAGMVFFQFYLFPHLTALENVMGPIRVRGVKKEEAQKAKAL 120

```

```

Query: 118 LERLGLKDHALVYPFSLSGGQKQVALARAMMIDPQIIGYDEPTSALDPELRQEVEKLIL 177
      L ++GL + A YP LSGGQ+QVVA+ARA+ + P+++ +DEPTSALDPELR EV K++
Sbjct: 121 LAKVGLAERAHHPSELSSGGQQQVVAIARALAVKPKMMLFDEPTSALDPELRHEVLKVMQ 180

```

Query: 178 QNRETGMTQIVVTHDLQFAESISDTILKIN 207

-1696-

E GMT ++VTH++ FAE ++ ++ I+  
 Sbjet: 181 DLAEGMTMVIVTHEIGFAEKVASRLIFID 210

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4703> which encodes the amino acid sequence <SEQ ID 4704>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2146(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 164/209 (78%), Positives = 183/209 (87%)

Query: 1 MLELKNISKCYGQKEIFKDFNLTVEEGKILSLVGPSGGGKTTLLRMLAGLEKIDSGTIVH 60  
 MLELKNISK +GQK IF FNLTV++G++LSLVGPS GGKTTLLRMLAGLE IDSG + +  
 Sbjet: 1 MLELKNISKQFGQKTIFDGFNLTVDGEVLSLVGPSSGGKTTLLRMLAGLESIDSGQVVFY 60

Query: 61 DGKEVSVDHLETLNLLGFVFQDFQLFPHLTVLDNLILSPVKTMGLSKELAKEKALVLLER 120  
 +G++V +DHLE NLLGFVFQDFQLFPHLTVLDNL LSP TMG K AKEKAL LL R  
 Sbjet: 61 NGEVDGIDHLENRNLLGFVFQDFQLFPHLTVLDNLTLSPITITMGKQKADAKEKALDLLAR 120

Query: 121 LGLKDHALVYPFSLSGGQKQORVALARAMMIDPQIIGYDEPTSALDPELRQVEKLILQNR 180  
 LGLK+HA VYP+SLSGGQKQORVALARAMMIDPQIIGYDEPTSALDPELRQ VE LI+QNR  
 Sbjet: 121 LGLKEHAQVYPFSLSGGQKQORVALARAMMIDPQIIGYDEPTSALDPELRQTVEALIVQNR 180

Query: 181 ETGMTQIVVTHDLQFAESISDTILKINPK 209  
 E G+TQIVVTHDL FAE+ISD I+++NPK  
 Sbjet: 181 EMGITQIVVTHDLVFAEASIDRIIRVNPK 209

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1528

A DNA sequence (GBSx1619) was identified in *S.agalactiae* <SEQ ID 4705> which encodes the amino acid sequence <SEQ ID 4706>. This protein is predicted to be amino acid ABC transporter, permease protein (glnP). Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -8.12 Transmembrane 102 - 118 ( 96 - 120)

----- Final Results -----

bacterial membrane --- Certainty=0.4248(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9341> which encodes amino acid sequence <SEQ ID 9342> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA98402 GB:AP002545 ABC amino acid transporter permease  
 [Chlamydophila pneumoniae J138]  
 Identities = 55/127 (43%), Positives = 83/127 (65%), Gaps = 1/127 (0%)

Query: 3 AAIIAFTMNYAAYFAEIFRGGIESIPKQGYEAAKVLKFSKFQTVWYIVLPQVFKIVLPSV 62  
 A IIA +MN AAY AE RGGI S+ GQ+E+A VL + K+Q YI+ PQVFK +LPS+

-1697-

Sbjct: 89 AGIIALSMSNAAYLAENIRGGINSLSIGQWESAMVLGYKKYQIFVYIIPQVFNILPSL 148

Query: 63 FNETITLVKDSSLVYILGVGDLLLESKTAANRDATLAPMF-IAGGIYLLIGLLTILSKQ 121  
NE ++L+K+SS++ ++GV +L +K +R+ M+ I G+Y L+ + +S+

5 Sbjct: 149 TNEFVSLIKESSILMVVGVPETKVTKDIVSRELNPMEYLIAGLYFLMTSSFSICISRL 208

Query: 122 VEKRFNY 128  
EKR +Y

10 Sbjct: 209 SEKRRSY 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4707> which encodes the amino acid sequence <SEQ ID 4708>. Analysis of this protein sequence reveals the following:

Possible\_site: 34

>>> Seems to have no N-terminal signal sequence

15 INTEGRAL Likelihood = -11.57 Transmembrane 21 - 37 ( 7 - 44)  
INTEGRAL Likelihood = -10.93 Transmembrane 185 - 201 ( 178 - 206)  
INTEGRAL Likelihood = -3.29 Transmembrane 63 - 79 ( 62 - 81)

----- Final Results -----

20 bacterial membrane --- Certainty=0.5628(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:BAB05181 GB:AP001512 ABC transporter (permease) [Bacillus halodurans]  
Identities = 86/206 (41%), Positives = 126/206 (60%), Gaps = 1/206 (0%)

Query: 4 IQQVLPSLLDGLVTLQVFFVILSIPLGAILAFLMKIPFKPLQWFLTLVWMMRGTP 63  
IQ +P +L+G VTLQ + ++ + LG +LA ++ +WF Y + RGTPL

30 Sbjct: 8 IQPFMPFMLEGVVWTLQFVSVSLFLGLVGLGIVLAIFKISKYRLFRWFADFYTSIFRGTP 67

Query: 64 LLQLIFFYYVLPVSGISFDRMPAAILAFTLNAAAYFAEIFRGGIEAIPKGQYEA AKVLKL 123  
+LQL+ Y LP G+ + AA LAF LN AAY +EI R GI+A+ KGQ EAA+ L +

35 Sbjct: 68 ILQLLMIYALPQFGVDISQFQAFLAFLNSAAYVSEIIRAGIQAVDKGQREAAEALGI 127

Query: 124 KPLQTIRYIILPQVFKIVLPSVFNEVINLVKDSSLVYVILGVGDLL-LASKTAANRDATLA 182  
+ IILPQ + +Lp++FNE INL K+S++V V+GV DL+ A T+A L

Sbjct: 128 PYRPMMLRIILPQAMRNILPALFNEFINLTKEAIVSVIGVTDLMRRAQITS AETYLE 187

40 Query: 183 PMFIAGLIYLLIGLVTTIISKQVEKR 208  
P+ GLIY +L+ +T+I + +E+R

Sbjct: 188 PLLFVGLIYYVLMGLTVIGRLERR 213

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 112/130 (86%), Positives = 121/130 (92%)

Query: 1 MPAAIIAFTMNAAAYFAEIFRGGIESIPKGQYEA AKVLKFSKFQTVWYIVLPQVFKIVLP 60  
MPAAI+AFT+NYAAAYFAEIFRGGIE+IPKGQYEA AKVLK QT+ YI+LPQVFKIVLP

50 Sbjct: 84 MPAAIIAFTLNAAAYFAEIFRGGIEAIPKGQYEA AKVLKLPQTIRYIILPQVFKIVLP 143

Query: 61 SVFNETITLVKDSSLVYILGVGDLLLESKTAANRDATLAPMFIAGGIYLLIGLLTILSK 120  
SVFNE I LVKDSSLVY+LGVGDLLL SKTAANRDATLAPMFIAG IYLLIGL+TI+SK

Sbjct: 144 SVFNEVINLVKDSSLVYVILGVGDLLLASKTAANRDATLAPMFIAGLIYLLIGLVTTIISK 203

55 Query: 121 QVEKRFNYYK 130  
QVEKRFNYY+

Sbjct: 204 QVEKRFNYYQ 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1698-

**Example 1529**

A DNA sequence (GBSx1620) was identified in *S.agalactiae* <SEQ ID 4709> which encodes the amino acid sequence <SEQ ID 4710>. This protein is predicted to be minidiscs. Analysis of this protein sequence reveals the following:

```

5   Possible site: 61
    >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.66    Transmembrane  44 - 60 ( 39 - 66)
      INTEGRAL    Likelihood = -7.96    Transmembrane 129 - 145 ( 123 - 147)
10   INTEGRAL    Likelihood = -5.15    Transmembrane  13 - 29 (  9 - 33)
      INTEGRAL    Likelihood = -2.39    Transmembrane  94 - 110 ( 94 - 110)

    ----- Final Results -----
          bacterial membrane --- Certainty=0.4864(Affirmative) < succ>
          bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:AAF49688 GB:AE003532 mnd gene product [Drosophila melanogaster]
    Identities = 48/145 (33%), Positives = 78/145 (53%), Gaps = 8/145 (5%)

20   Query: 7   IKQTYGLMTTIAMIVGVVIGSGIYFKVDDILKFTGGDVFLGMVILVLGSFSIVFGSL SIS 66
      +K+  GL+  +A+IVGV++GSGI+      +LKF+ G +   +++ VL      + G+L +
    Sbjct: 39   LKKQIGLLDGVAIIVGVIVGSGIFVSPKGVLKFS-GSIGQSLIVVVLSGVLSMVGALCYA 97

25   Query: 67   ELAIRTSESGGIFSYYEKYVSPALAAATLGLFASFLYL-PTLTAIVSWVAIFYTLGE---- 121
      EL      +SGG ++Y      P L A L L + + L L PT AI +   A Y L
    Sbjct: 98   ELGTMIPKSGGDYAYIGTAFGP-LPAFLYLWVALLILVPTGNATTALTFAIYLLKPFWPS 156

30   Query: 122' -SSSLESQIILA AVYILALSLMNIF 145
      + +E+ +LAA I L+L+N +
    Sbjct: 157   CDAPIEAVQLLAAAMICVLTLINCY 181

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

**Example 1530**

A DNA sequence (GBSx1621) was identified in *S.agalactiae* <SEQ ID 4711> which encodes the amino acid sequence <SEQ ID 4712>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 21
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
45   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1699-

**Example 1531**

A DNA sequence (GBSx1622) was identified in *S.agalactiae* <SEQ ID 4713> which encodes the amino acid sequence <SEQ ID 4714>. This protein is predicted to be TRK potassium uptake system protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 27
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.06    Transmembrane  232 - 248 ( 232 - 248)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8835> which encodes amino acid sequence <SEQ ID 8836> was also identified. Analysis of this protein sequence reveals the following:

```

15  Lipop: Possible site: -1    Crend: 5
    McG: Discrim Score:      -4.65
    GvH: Signal Score (-7.5): -3.64
    Possible site: 27

20  >>> Seems to have no N-terminal signal sequence
    ALOM program  count: 1 value: -0.06 threshold:  0.0
      INTEGRAL    Likelihood = -0.06    Transmembrane  228 - 244 ( 228 - 244)
      PERIPHERAL Likelihood =  1.27      428
    modified ALOM score:  0.51

25  *** Reasoning Step: 3

    ----- Final Results -----
30  bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
    bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

35  >GP:AAB90401 GB:AE001046 TRK potassium uptake system protein
    (trkA-2) [Archaeoglobus fulgidus]
    Identities = 136/446 (30%), Positives = 238/446 (52%), Gaps = 12/446 (2%)

Query: 5   MRIIVVGGGKVG TALCRSLVAEKHDVVLIEKKENVLKRVTKQHDIMGIVGNGANYKILEQ 64
Sbjct: 1   MRIVIAGAGEVGYHLAMSL-APNHDVIIIEKDVSRFERVSEL-DVVAINGNAANMKVLRD 58

Query: 65  AEVKNCIDIFIAITDRDEVNMMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFSSV 124
Sbjct: 59  AGVERADVFLAVTGNDEVNLLSGLAAKKGAKNVIVRVENPEYVDRPIVKEHPLGYDVL 118

45  Query: 125 NPELLAAQYIANTIEFPNATSV EHFANGRVMLMEFKILEGNKLC HTSMSQIRKKFGNIVI 184
Sbjct: 119 CPQLSLAQEAARLIGIPGAIEVVTFSGGKVEMIELQVMEGSKADGKAIADLYLP-QNVVI 177

50  Query: 185 CAIERDGKLIIPDG DATIQVKDKIFVTGNRIEMILFHN YVKNKVKNLMVIGAGRIAYYL 244
Sbjct: 178 ASIYRNHGIEIPRGDTVLRAGDRVAIVSKTEDVEMLKGIFGPPVTRRVTFGAGTIGSYT 237

55  Query: 245 LNILKNTNTHVKLV ELNQEQAEYFSQEFNPVPVHGDGTAKNILLEESVTSF DAVATLTG 304
Sbjct: 238 AKILAKGMTSVKLI ESSMERCEALS GELEGVRIVCGDATDIEFLIEEEIGKSDAVLAATE 297

Query: 305  VDEENIITSMFLESIGIPKNITKVNRTSLLEIIDDQQLSSIITPKRIAVDHVMHFVRGRV 364
Sbjct: 298 SDEKNLLISL LSKNLGARIAIAKVEKREYVKLF EAVGV DVALNPRSVTYNEVSKLLR--- 354

60  Query: 365 NAQDSNLEAMHHIANDRIETLQFEIKETSKLANRSLASLKLKQNILIAAIRNNKTIFPT 424

```

-1700-

          +E + I + + + ++L ++L L L ++ +I AI+R N+ + P  
 Sbjct: 355 ---TMRITLAEIEGTAVVEV---VVRNTRLVKGKALKDLPLPKDAIIGAIVRGNECLIPR 408

Query: 425 GEDVLTVGDRIVVITLLKNITRTSDM 450

G+ + DR++V I + ++

Sbjct: 409 GDTTIEYEDRLLVFAKWDEIEKIEEI 434

Identities = 48/212 (22%), Positives = 99/212 (46%), Gaps = 15/212 (7%)

Query: 3 VKMRIIVGGGKVGKTALCRSLVAEKHDVVLEIEKKENVLKRVTQHDIMGIV-GNGANYKI 61

V R+ + G G +G+ + L V LIE + + + + + IV G+ + +

Sbjct: 221 VTRRVTFIFGAGTIGSYTAKILAKGMTSVKLISSMERCEALSSELEGVRIVCGDATDIEF 280

Query: 62 LEQAEVKNCDIFIAITDRDEVNMISAVLAKKMGAKETVVRMRNPEYSNPYFKDKNFLGFS 121

L + E+ D +A T+ DE N++ ++L+K +GA+ + ++ EY + +G

Sbjct: 281 LIEEEIGKSDAVLAATESDEKNLLISLLSKNLGARIAIAKVEKREYVKLF----EAVGVD 336

Query: 122 SVVNPELLAAQYIA---NTIEFPNATSVVEHFANGRVMLMEFKILEGNKLCHTSMSQIRKK 178

+NP + ++ T+ +E A V++ +++ G L + +

Sbjct: 337 VALNPRSVTYNEVSKLLRTMRIETLAEIEGTAVVEVVVRNTRLV-GKALKDLPLPK---- 391

Query: 179 FGNIVICAIERDGLIIPDGDATIQVKDKIFV 210

+ +I AI R + +IP GD TI+ +D++ V

Sbjct: 392 --DAIIGAIVRGNECLIPRGDTTIEYEDRLLV 421

There is also homology to SEQ ID 4716.

SEQ ID 8836 (GBS384) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 2; MW 53kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 6; MW 78kDa).

The GBS384-GST fusion product was purified (Figure 212, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 279), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1532

A DNA sequence (GBSx1623) was identified in *S.agalactiae* <SEQ ID 4717> which encodes the amino acid sequence <SEQ ID 4718>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4948(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1701-

**Example 1533**

A DNA sequence (GBSx1624) was identified in *S.agalactiae* <SEQ ID 4719> which encodes the amino acid sequence <SEQ ID 4720>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood ==-12.58    Transmembrane    37 - 53 ( 33 - 61)
    INTEGRAL    Likelihood ==-11.57    Transmembrane    183 - 199 ( 179 - 214)
    INTEGRAL    Likelihood ==-10.03    Transmembrane    397 - 413 ( 392 - 424)
10  INTEGRAL    Likelihood = -6.79     Transmembrane    14 - 30 ( 5 - 31)
    INTEGRAL    Likelihood = -6.42     Transmembrane    71 - 87 ( 69 - 93)
    INTEGRAL    Likelihood = -4.99     Transmembrane    278 - 294 ( 274 - 295)
    INTEGRAL    Likelihood = -4.19     Transmembrane    133 - 149 ( 132 - 152)
    INTEGRAL    Likelihood = -4.09     Transmembrane    327 - 343 ( 324 - 344)
    INTEGRAL    Likelihood = -2.44     Transmembrane    236 - 252 ( 234 - 252)
15  INTEGRAL    Likelihood = -0.59     Transmembrane    456 - 472 ( 456 - 472)

----- Final Results -----
    bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10065> which encodes amino acid sequence <SEQ ID 10066> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

25  >GP:AAB90400 GB:AE001046 TRK potassium uptake system protein (trkH)
    [Archaeoglobus fulgidus]
    Identities = 166/480 (34%), Positives = 262/480 (54%), Gaps = 10/480 (2%)

30  Query: 1  MNKSMIRFLLSKLLIEAALLAIPLTVGLIYREP-QSVMMSIVITMIILILGLLGSFLK 59
    MN + +L KLL++ + +PL ++ EP ++ +++++ +LG G +
    Sbjct: 1  MNLRLTASILGKLLMLFSFSFILPIAAHVFEFPYHPFLIPAALSLLVGAVLGY-GIKTE 59

    Query: 60  PKNYHIYTKEGMLIVALCWILWSFFGALPFVISGQIPNIIDAFFEVSSGFTTTGATILDD 119
    + + KE IVAL W+ S FG++P++I G P +DAFFE SGFTTTGA++L
35  Sbjct: 60  SEFDSL RHKESFAIVALIWLFMISIFGSIPYIIFGISP--VDAFFESMSGFTTTGASVLTP 117

    Query: 120  VSVLSPALLFWRSFTHLIGGMGVLVFALAIMENSKNSHLEVMRAEVPGPVFGKVSKLKK 179
    L +LL WRS T IGGMG++V LAI N + +AE PG K+ +++
40  Sbjct: 118  EE-LPKSLLLWRSLTQWIGGMGIIVLFLAIFPNVAKRSTVLFQAEYPGVSLSKLPRI 176

    Query: 180  TAQILYLLYLLMFVAVFVILYFAGMPFFDSIIIAMGTAGTGGFAVYNDIAHYNSPLITN 239
    TA LY +YLL+ +LY G+ FD+I T TGG++ +++SIA + +
    Sbjct: 177  TALSLYKVYLLLTIAEVALLYALGLSLFDAINHTFTTLSTGGYSTHSESIATFFKDVRVEA 236

45  Query: 240  LVSIGMLIFGVNFNLYYLLLRKIKAFFGDEELKTYLRIVAIATFMIALNVIGMYDNFRQ 299
    +V+ + G NF L Y LL K F + E + Y+ +A+A+ +IA + Y F +
    Sbjct: 237  VVAFFAFLGANFALYIFLLSGK-PVIFRNTFRAYVCFALASVVIAAVNLD RYSIF-E 294

    Query: 300  GLEHIFFEVSAIITTTGFGVTDITRWPLFSQVILLFLMFIGGSAGSTAGGFKVMRSLILA 359
    L + F+ +I+TTTGF D W +++IL+ LMFIGGS+GST GG KV+R +L
50  Sbjct: 295  SLRYSIFQAVSIMTTTGFTTADFDAWSDSAKLILVVL MFIGGSSGSTGGGIKVIRIYLLI 354

    Query: 360  KIARNQVLSTLYPNRVMSLHINKSVLKDNTQHGVLYLYLAIFMALVLVLTLDNDNFL 419
    K A +Q+L P V ++ + K + + +Y+ IF +++L D +
55  Sbjct: 355  KYAVHQILRAAEPRTVRVAKFEGRAIKKEILDDIAAFFVLYILIFAVSSILVSLSGYDIV 414

    Query: 420  VVISAAASCNNIGP---LLGSNETFSFFSPFSKLLLSFAMIAGRLEIYPVLLMFIPK 476
    ISA A+ N+GP L G+ E ++ F +K+LL+ M GRLEI+ V+ +FIP W
60  Sbjct: 415  TSISATAATLGNVGPGLAGAAENYASFPSLT KILLAVNMWIGRLEIFTVVSLFIPTFW 474

```

No corresponding DNA sequence was identified in *S.pyogenes*.

-1702-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1534

A DNA sequence (GBSx1625) was identified in *S.agalactiae* <SEQ ID 4721> which encodes the amino acid sequence <SEQ ID 4722>. Analysis of this protein sequence reveals the following:

Possible site: 22  
>>> Seems to have no N-terminal signal sequence (or aa 1-20)

----- Final Results -----

10                   bacterial cytoplasm --- Certainty=0.2870(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15       >GP:AAD36530 GB:AE001797 conserved hypothetical protein  
          [Thermotoga maritima]  
          Identities = 43/75 (57%), Positives = 57/75 (75%), Gaps = 1/75 (1%)  
  
Query: 2   LKSFLIFLVRFYQKNISPAPFASCRYRPTCSTYMIEAIQKHG-LKGVLMGIARILRCHPL 60  
          +K LI L+RFYQ+ ISP P +CR+ PTCS Y I+A++KHG LKG +G+ RILRC+PL  
20       Sbjct: 1 MKKLLIMLIRFYQRYISPLKPPTCRFTPTCSNYFIQALEKHGLLKGTFLGLRRILRCNPL 60  
  
Query: 61 AHGGNDPVPDHFSLR 75  
          + GG DPVP+ FS +  
25       Sbjct: 61 SKGGYDFVPEEFSFK 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4723> which encodes the amino acid sequence <SEQ ID 4724>. Analysis of this protein sequence reveals the following:

30       Possible site: 38  
          >>> Seems to have no N-terminal signal sequence  
  
          ----- Final Results -----  
                  bacterial cytoplasm --- Certainty=0.3639(Affirmative) < succ>  
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
35               bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 53/78 (67%), Positives = 60/78 (75%)  
  
40       Query: 1 MLKSFLIFLVRFYQKNISPAPFASCRYRPTCSTYMIEAIQKHGLKGVLMGIARILRCHPL 60  
          M+K LI V+ YQK ISP P SCRY+PTCS YM+ AI+KHG KG+LMGIARILRCHP  
          Sbjct: 1 MMKLLIVSVKAYQKYISPLSPFSCRYKPTCSAYMLTAIEKHGKTKGILMGIARILRCHPF 60  
  
Query: 61 AHGGNDPVPDHFSLRRNK 78  
          GG DPVP+ FSL RNK  
45       Sbjct: 61 VAGGVDPVPEDFSLMRNK 78

SEQ ID 4722 (GBS233) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 3; MW 35.6kDa).

50       The GBS233-GST fusion product was purified (Figure 207, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 280), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1703-

**Example 1535**

A DNA sequence (GBSx1626) was identified in *S.agalactiae* <SEQ ID 4725> which encodes the amino acid sequence <SEQ ID 4726>. This protein is predicted to be ribosomal large subunit pseudouridine synthase B (rluB). Analysis of this protein sequence reveals the following:

```

5   Possible site: 18
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10          bacterial cytoplasm --- Certainty=0.2957(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:BAB05295 GB:AP001512 pseudouridylate synthase [Bacillus halodurans]
    Identities = 130/239 (54%), Positives = 175/239 (72%), Gaps = 2/239 (0%)

    Query: 2   RINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAG-DLVEIEGSPYNEEKV 60
              R+ K IA AGIASRRKAE+LI +G V +NGQVV EL +V D +E+EG P+ EE V
    Sbjct: 3   RLQKVIAQAGIASRRKAEQLILEGKVKVNGQVVKELGIKVNPNQDDIEVEGVFVEKEEVP 62

20   Query: 61  YYLLNKPRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKM 120
              Y+LL KP GVISSV DDKGRK V D L ++++R+YPVGRLD+DT+GLL+LTNDG+F + +
    Sbjct: 63  YFLLYKPTGVISSVKDDKGRKVVTDFL-EIEQRVYPVGRLDYDTSGLLLLTNDGEFANLL 121

25   Query: 121 IHRNEIDKVYLARVKGIATKENLRPLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVVELT 180
              +HPR++I+KVY+A+VKGI T++ L+ L RGV ++ T PA+ ++ VD K ++V+LT
    Sbjct: 122 MHRPHKIEKVYVAKVKGIPTRDQLKLLARGVKLEDGPTAPAKVKMLSVDRRKQTAIVKLT 181

30   Query: 181 IHEGRNHQVKKMFQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKESQLHNAAINK 239
              IHEGRN QV++MFE +G V KL R QF LDL+G+ PG+ R L E+ L A+ K
    Sbjct: 182 IHEGRNRQVRRMFETIGCEVMKLRQFAFLDLSGMNPGDVRPLKPIEVKHLRELAVTK 240

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4727> which encodes the amino acid sequence <SEQ ID 4728>. Analysis of this protein sequence reveals the following:

```

35   Possible site: 18
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
40          bacterial cytoplasm --- Certainty=0.1587(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

45   Identities = 210/239 (87%), Positives = 228/239 (94%)

    Query: 1   MRINKYIAHAGIASRRKAEELIKQGMVTINGQVVNELATQVKAGDLVEIEGSPYNEEKV 60
              MRINKYIAHAGIASRRKAEELIKQ+VT+NGQV+ +LAT VK+GD+VEIEGSPYNEEKV
    Sbjct: 9   MRINKYIAHAGIASRRKAEELIKQLVTLNGQVITDLATTVKSGDVVEIEGSPYNEEKV 68

50   Query: 61  YYLLNKPRGVISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTTGLLILTNDGDFTDKM 120
              YYLLNKPRG ISSVSDDKGRKTIV+DLLPQVKERIYPVGRLDWDT+G+LILTNDGDFTD M
    Sbjct: 69  YYLLNKPRGAISSVSDDKGRKTVIDLLPQVKERIYPVGRLDWDTSGVLILTNDGDFTDTM 128

55   Query: 121 IHRNEIDKVYLARVKGIATKENLRPLTRGVVIDGKKTKPARYTIIKVDHEKNRSVVVELT 180
              IHRNEIDKVYLARVKGIATKENLRPLTRG+VIDGKKTKPARY I++V+ +K+RS+VELT
    Sbjct: 129 IHRNEIDKVYLARVKGIATKENLRPLTRGIVIDGKKTKPARYNIVRVEADKRSRIVELT 188

    Query: 181 IHEGRNHQVKKMFQVGLLVDKLSRTQFGTLDLTGLRPGEARRLNKKESQLHNAAINK 239
              IHEGRNHQVKKMF VGLLVDKLSRT+FGT+DL GLRPGEARRLNKKESQLHN A K
60   Sbjct: 189 IHEGRNHQVKKMFESVGLLVDKLSRTRFGTVDLKGRLPGEARRLNKKESQLHNLANTK 247

```

-1704-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1536

A DNA sequence (GBSx1627) was identified in *S.agalactiae* <SEQ ID 4729> which encodes the amino acid sequence <SEQ ID 4730>. Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10           bacterial cytoplasm --- Certainty=0.1476(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15   >GP:BAB05280 GB:AP001512 unknown conserved protein [Bacillus halodurans]  
     Identities = 75/180 (41%), Positives = 107/180 (58%), Gaps = 10/180 (5%)  
  
     Query: 6   SIEALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLLLETAQTYKL 65  
              +IE +LFV G++G++L ++ +LL L+    +   LE+L   Y D+   L + E A ++L  
20   Sbjct: 9   AIEGILFVRGDEGVTLEELCDLLELSTDVVYAAL EELRLSYT-DEAGRLRIEEVAHAFL 67  
  
     Query: 66   VTKDSYMTLLRDYAKAPINQSLSRASLEVL SIIAYKQPITRIEIDDIRGVNSSGAITRLI 125  
              TK           + A + +   LS+A+LE L+IIAY+QPITRIE+D++RGV S AI L  
25   Sbjct: 68   STKPELAPYFKKLALSTLQSGLSQAALETLAI IAYRQPITRIEVDVRGVKSEKAIQTLT 127  
  
     Query: 126 AFGLIKEAGKKEVLGRPNLYETTNYFLDYMGINQLDDL-----IDASSIELVDEEVSLF 179  
              + LIKE G+ + GRP LY TT FLD+ G+ L +L       ID SSI   EE LF  
30   Sbjct: 128 SRLLIKEVGRAQGTGRPILYGTTPQFLDHFGLKSLKELPPLPEDIDESSI---GEEADLF 184

30   A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4731> which encodes the amino acid sequence <SEQ ID 4732>. Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35           bacterial cytoplasm --- Certainty=0.1062(Affirmative) < succ>  
             bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40   An alignment of the GAS and GBS proteins is shown below.

Identities = 130/179 (72%), Positives = 159/179 (88%)

     Query: 1   MTYLGSI EALLFVAGEDGLSLRQMAELLSLTPSALIQQLEKLAKRYEEDDDSSLLLLLETA 60  
              MTYL IEALLFVAGE+GLSLR +A +LSLTP+AL QQLEKL+++YE+D SSL L+ETA  
45   Sbjct: 1   MTYLSQIEALLFVAGEEGLSLRHLASMLS LTPTALQQLEKLSQKYEKDQHSSLCIETA 60  
  
     Query: 61   QTYKLVTKDSYMTLLRDYAKAPINQSLSRASLEVL SIIAYKQPITRIEIDDIRGVNSSGA 120  
              TY+LVTK+ + LLR YAK P+NQSLSRASLEVL SI+AYKQPITRIEIDDIRGVNSSGA  
50   Sbjct: 61   NTYRLVTKEGFAELLRAYAKTPMNQSLSRASLEVL SIVAYKQPITRIEIDDIRGVNSSGA 120  
  
     Query: 121 ITRLIAFGLIKEAGKKEVLGRPNLYETTNYFLDYMGINQLDDLIDASSIELVDEEVSLF 179  
              +++L+AF LI+EAGKK+V+GRP+LY TT+YFLDYMGIN LD+LI+ S++E DEE++LF  
     Sbjct: 121 LSKLLAFDLIREAGKKDVGPHLYATTDYFLDYMGINHLDELIEVSAVEPADEEIALF 179

55   Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1705-

**Example 1537**

A DNA sequence (GBSx1628) was identified in *S.agalactiae* <SEQ ID 4733> which encodes the amino acid sequence <SEQ ID 4734>. Analysis of this protein sequence reveals the following:

Possible site: 47

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1012(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 89/246 (36%), Positives = 145/246 (58%), Gaps = 19/246 (7%)

Query: 3 IKLKDFEGPLDLLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMASQ 62  
+K+ FEGPLDLLLHL+++ E+DIYD+P+ ++ EQYL Y+ T++ + L++A EY++MA+  
Sbjct: 6 VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63 LMLIKSRNLLPK----VVESNPI-EDDPEMELLSQLEFYRRFKVLSEELANQHQERAKYF 117  
L+ IKSR LLPK + E + E+DP EL+ +L EYR++K +++L + +ER K F  
Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLEEDPREELIEKLEIYRKYKDAADLKEREERQKSF 125

Query: 118 SKPKQEVIFEDAILLHDKSVMDLFLTFSQMMSQKQKELSNS-----QTVIEKEDYRIED 171  
+KP ++ + +S L +T M+ QK L +T I ++D IE  
Sbjct: 126 TKPPSDL--SEYAKEVKQSEQKLSVTIVYDMIGAFQKVLKRKKINRPMETTITRQDIPIEA 183

Query: 172 MMIVIERHFNLLKKTT---LQEVFADCQTKSEMITLFLAMLELIKHLQITVEQDSNFSQV 228  
M I +LK + T ++F + K ++ FLA+LEL+K + +EQ+ NFS +  
Sbjct: 184 RMNEIVH--SLKSRGTRINFMDLF-PYEQKEHLVVTFLAVLELMKNQLVLIEQEHNFSDI 240

Query: 229 ILRKEE 234  
+ E

Sbjct: 241 YITGSE 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4735> which encodes the amino acid sequence <SEQ ID 4736>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -3.61 Transmembrane 199 - 215 ( 199 - 218)

----- Final Results -----

bacterial membrane --- Certainty=0.2444(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB14254 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]  
Identities = 86/239 (35%), Positives = 145/239 (59%), Gaps = 15/239 (6%)

Query: 3 IKLKDFEGPLDLLLHLVSKYKVDIYEVPIVEVIEQYLYNIETLQVMKLEVAGDYMLMASQ 62  
+K+ FEGPLDLLLHL+++ ++DIY++P+ ++ EQYL Y+ T++V++L++A +Y++MA+  
Sbjct: 6 VKIDTFEGPLDLLLHLINRLEIDIYDIPVAKITEQYLLYVHTMRVLELDIASEYLVMAAT 65

Query: 63 LMLIKSRRLLPKVVEHI-----EEDLEQDLLEKIEEYSRFAVSAQALAKQHDQRAK 115  
L+ IKSR LLPK E + EED ++L+EK+ EY ++K ++ L ++ ++R K +  
Sbjct: 66 LLSIKSRMLLPKQEEELFEDELLEEDPREELIEKLEIYRKYKDAADLKEREERQKSF 125

Query: 116 SKPKQELI-FEDAILQEDK----TVMDLFLAFSNIMAARAVLKNHNTVIERDDYKIEDM 170  
+KP +L + + Q ++ TV D+ AF ++ K+ + + T I R D IE  
Sbjct: 126 TKPPSDLSEYAKEVKQSEQKLSVTIVYDMIGAFQKVLKRKK-INRPMETTITRQDIPIEA 184

-1706-

Query: 171 MASIKQRLEKENV-IRLSAIFEECQTLNEVISIFLASLELIKLVHVFVEQLSNFGAIL 228  
 M I L+ I +F Q + V++ FLA LEL+K +V +EQ NF I +  
 Sbjct: 185 MNEIVHSLKSRGTRINFMDLFPYEQKEHLVVT-FLAVLELMKNQLVLIEQEHNFSDIYI 242

An alignment of the GAS and GBS proteins is shown below.

Identities = 156/235 (66%), Positives = 191/235 (80%), Gaps = 2/235 (0%)

Query: 1 MDIKLKDFEGPLDLLHLVSKYEVDIYDVPIVEVIEQYLAYIATLQAMRLEVAGEYMLMA 60  
 MDIKLKDFEGPLDLLHLVS+Y+VDIY+VPIVEVIEQYL YI TLQ M+LEVAG+YMLMA  
 Sbjct: 1 MDIKLKDFEGPLDLLHLVSQYKVDIYEVPIVEVIEQYLYNIETLQVMKLEVAGDYMLMA 60

Query: 61 SQLMLIKSRNLLPKVVESNPIEDDPMEMLLSQLEEYRRFKVLSEELANQHQERAKYFSKP 120  
 SQLMLIKSR LLPKVVE IE+D E +LL ++EY RFK +S+ LA QH +RAK++SKP  
 Sbjct: 61 SQLMLIKSRLLPKVVEH--IEEDLEQDLLEKIEEYSRFKAVSQALAKQHDQRAKWYSKP 118

Query: 121 KQEVIFEDAILLHDKSVMDFLTFSQMSQKQKELSNSQTVIEKEDYRIEDMMIVIERHF 180  
 KQE+IFEDAIL DK+VMDLFL FS +M+ K+ L N+ TVIE++DY+IEDMM I++  
 Sbjct: 119 KQELIFEDAILQEDKTVMDLFLAFSNIMAAKRAVLKNNHTVIERDDYKIEDMMASIKQRL 178

Query: 181 NLKKKTTLQEVFADCQTKSEMITLFLAMLELIKLVHQTVEQDSNFSQVILRKEEK 235  
 + L +F +CQT +E+I++FLA LELIKLH + VEQ SNF +ILRKE+K  
 Sbjct: 179 EKENVIRLSAIFEECQTLNEVISIFLASLELIKLVHVFVEQLSNFGAILRKEKK 233

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1538

A DNA sequence (GBSx1629) was identified in *S.agalactiae* <SEQ ID 4737> which encodes the amino acid sequence <SEQ ID 4738>. This protein is predicted to be pXO1-18. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -4.14 Transmembrane 128 - 144 ( 127 - 145)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2657(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05248 GB:AP001512 integrase/recombinase [Bacillus halodurans]  
 Identities = 67/271 (24%), Positives = 117/271 (42%), Gaps = 35/271 (12%)

Query: 11 LKTMINDINNFIKSK---LSINSRKSYPHYDLKQFYKII-----GGHVNSEKLALY 58  
 ++T+ N++ F+ +K LS N+ +SY DLKQ+ + + ++ E + Y  
 Sbjct: 1 METVNNNLQQFLHFQKVERGLSNNTIQSYGRDLKQYIQYVERVEEIRSARNITRETILHY 60

Query: 59 QQSLSEFKL--TARKRKL SAVNQFLFFLYNRGTLKEFYRL-----QETEKITLAQTKSQI 111  
 L E T+ R ++A+ F FL + + T+++ A T ++  
 Sbjct: 61 LYHLREQGRAETS IARVAVAIRSFHQFLLEKLSDSPTVHVEIPKATKRLPKALTIEEV 120

Query: 112 MDLSNFYQDTPSGRLIALLL--SLGLTPAEIANLKKADFDTTFNLS-IEKSQMKRI 168  
 L N Q D S R A+L L + G+ +E+ L +D + + + K +RI  
 Sbjct: 121 EALLNSPQGRDPFSLRNKAMLELLYATGMVSELIGLTLSDIHLMSGFVRCCLGKGNKRI 180

Query: 169 LKLPEDLLPFLLESLEEDG-----DLVF-EHNGKPYSRQWFFNQLTDFLNEKN-E 216  
 + + + + +ES +G D VF H+G+P SRQ F+ L N +  
 Sbjct: 181 IPIGQ-VATEAVESYLANGRGKLMKKQSHDHVVFVNHGRPLSRQGFWMKLQAKNVNID 239

Query: 217 QQLTAQLLREQFILKQKENGKTMTELSRLLG 247

-1707-

+ LT LR F ENG + + +LG  
 Sbjct: 240 KPLTPHTLRHSFATHLLENGADLRVQEMLG 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4739> which encodes the amino acid sequence <SEQ ID 4740>. Analysis of this protein sequence reveals the following:

Possible site: 20  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -0.90 Transmembrane 111 - 127 ( 110 - 127)

----- Final Results -----  
 bacterial membrane --- Certainty=0.1362(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/243 (48%), Positives = 167/243 (68%), Gaps = 1/243 (0%)

Query: 18 INNFISSKKLSLNSRKSYHYDLKQFYKIIGGHVNSEKLALYQQSLSEFKLTARKRKLSAV 77  
 I FI SK LS NS+K+Y YDL+QF ++IG VN +KL LYQ S++ L+A+KRKLS  
 Sbjct: 5 IEPFIASKALSQNSQKAYRYDLQQFCQLIGERVNQDKLLLYQNSIANLSLSAKKRKLSA 64

Query: 78 NQFLFFLYNNGTLKEFYRLQETEKITLAQTK-SQIMDLNFIYQDTPDPSGRLIALLLISL 136  
 NQFL++LY L ++RL +T K+ + + + I++ FYQ T + G+LI+LLIL L  
 Sbjct: 65 NQFLYYLYQIKYLSYFRLTDTMKVMRTEKQQAIIINTDIFYQKTPFVWGQLISLLILEL 124

Query: 137 GLTPAEIANLKKADFDFTFNILSIEKSQMKRILKLPEDLLPFLLESLEEDGDLDVFEHNGK 196  
 GLTP+E+A ++ A+ D F +L+++ + R+L L + L+PFL + L +FEH G  
 Sbjct: 125 GLTPSEVAGIEVANLDLNFQMLTLKTKGVRVPLSQLIPFLEQQLVGKEVYLFEHRGI 184

Query: 197 PYSRQWFFNQLTDFLNEKNEQQLTAQLLREQFILKQKENGKTMTELSRLLGLKTPITLER 256  
 P+SRQWFFN L F+ + LTAQ LREQFILK+K GK++ ELS +LGLK+P+TLE+  
 Sbjct: 185 PFSRQWFFNHLKTFVRSIGYGLTAQKLREQFILKEKLAGKSIIEISDILGLKSPMTLEK 244

Query: 257 YYR 259  
 YY+  
 Sbjct: 245 YYK 247

SEQ ID 4738 (GBS383) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 68 (lane 7; MW 32kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 5; MW 57.1kDa).

The GBS383-GST fusion product was purified (Figure 212, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 308), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1539

A DNA sequence (GBSx1630) was identified in *S.agalactiae* <SEQ ID 4741> which encodes the amino acid sequence <SEQ ID 4742>. Analysis of this protein sequence reveals the following:

Possible site: 21  
 >>> Seems to have no N-terminal signal sequence  
 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2465(Affirmative) < succ>

-1708-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAB05201 GB:AP001512 unknown conserved protein in B. subtilis  
 [Bacillus halodurans]  
 Identities = 38/136 (27%), Positives = 73/136 (52%), Gaps = 1/136 (0%)  
 10 Query: 7 ESFLLNHLHDHYLIPAEDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGTISISDI 66  
 ++ + N L +IP E VA ++ +H +L+L +G++ +PV+ + K G IS S I  
 Sbjct: 7 QNIMDNDLKLVIPIFEKVAHVHLSNPLEHALLVLIKSGYTAIPVLDEHSKLHGVIKSLI 66  
 Query: 67 MGYQSKGQLTDWE-MAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISDQNDFR 125  
 + + + E +A + +++N +I I+ A+ + + + +PF+ ++ D F  
 15 Sbjct: 67 LDALLGVERIEMERLAHLVVKDVMNPEIPTIHHKASFSRALKVSIAPFICILDDGGSFL 126  
 Query: 126 GIITRKSILKAINSLL 141  
 GI+TR +IL IN L  
 Sbjct: 127 GILTRSTILSFINRQL 142  
 20

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4743> which encodes the amino acid sequence <SEQ ID 4744>. Analysis of this protein sequence reveals the following:

Possible site: 47  
 >>> Seems to have no N-terminal signal sequence  
 25 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.3539(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 30

An alignment of the GAS and GBS proteins is shown below.

Identities = 119/153 (77%), Positives = 137/153 (88%)  
 35 Query: 1 MIAKEFESFLLNHLHDHYLIPAEDVAIFVDTHNADHVMLLLASNGFSRVPVITKEKKYVGT 60  
 MIAKEFE+FL++HLD+YLIP +D+AIF+DTHNADHVMLLL SNGFSRVPVIT+EKKYVGT  
 Sbjct: 1 MIAKEFETFLMSHLDNYLIPEQDLAIFIDTHNADHVMLLLVSNNGFSRVPVITREKKYVGT 60  
 Query: 61 ISISDIMGYQSKGQLTDWEMAQTDIVEMVNTKIEPINEAATLTAIMHKIVDYPFLPVISD 120  
 ISISDIM YQSK QLTDWEM+QTDI EMVNTKIE I+ ++LT IMHK++D+PFLPV+  
 40 Sbjct: 61 ISISDIMYQSKRQLTDWEMSQTDIGEMVNTKIETISITSSLTAIMHKLIDFPFLPVVDR 120  
 Query: 121 QNDFRGIITRKSILKAINSLLHDFDTEYTTITPK 153  
 N F GIITRKSILKA+NSLLHDFD+YTI K  
 Sbjct: 121 ANRFVGIITRKSILKAVNSLLHDFDDEYTIKK 153  
 45

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1540

50 A DNA sequence (GBSx1631) was identified in *S.agalactiae* <SEQ ID 4745> which encodes the amino acid sequence <SEQ ID 4746>. Analysis of this protein sequence reveals the following:

Possible site: 14  
 >>> Seems to have no N-terminal signal sequence  
 55 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.4421(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1709-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA06785 GB:AP001517 unknown conserved protein [Bacillus halodurans]  
Identities = 55/169 (32%), Positives = 95/169 (55%), Gaps = 1/169 (0%)

5 Query: 5 KLIVMSDSHGDRDIVKDIKNHYLGKVD AIFHNGDSELPSSDPIWEGIHVVVGNC DYDSGY 64  
KL+++SDSHG D +K + + + +VDAI H GDSELP D EG+++V GNCD+ +  
Sbjct: 2 KLLILSDSHGWSDELKAVADKHRQEVDAIHC GDSELPDRDDRAEGMNIVRGNCDFGVDF 61

10 Query: 65 PEVLVTKIDNAVIVQTHGHLHQINFTWDKLDLLAQ QEDADICLYGHLHRADAWKNGKTIF 124  
PE + + + + THGHL+ + + L A++ A + +GH H A +++ +F  
Sbjct: 62 PEDFIKTVGDFNVYVTHGHLYNVKMSYVSLTYRAEEVGAQLVCFGHSHVATSFQENGIVF 121

15 Query: 125 INPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHQPYPNLTKELSR 173  
+NPGS+ PR E+ Y + + D ++ + + R +L + R  
Sbjct: 122 VNPGLRLPRNR-KEQTYCLAYVRDDQIELTFLDRDGHEVTDLQRTYLR 169

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4747> which encodes the amino acid sequence <SEQ ID 4748>. Analysis of this protein sequence reveals the following:

20 Possible site: 14  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3835(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 117/173 (67%), Positives = 143/173 (82%)

30 Query: 1 MAIRKLIVMSDSHGDRDIVKDIKNHYLGKVD AIFHNGDSELPSSDPIWEGIHVVVGNC DY 60  
MA + ++VMSDSHGDRDIV+ IK+ YLG+VDAIFHNGDSEL SSDPIW GI+VV GNCDY  
Sbjct: 1 MASKTIIVMSDSHGDRDIVQAIDKYLQGVDAIFHNGDSELNSSDPIWAGIYVVGNC DY 60

35 Query: 61 DSGYPEVLVTKIDNAVIVQTHGHLHQINFTWDKLDLLAQ QEDADICLYGHLHRADAWKNG 120  
D+GYP+ LVT++ I QTHGHL+ INFTWDKLD AQ+ ADICLYGHLHR AW+ G  
Sbjct: 61 DTGYPDRLVTLGTVTIAQTHGHLHYINFTWDKLDYFAQEVVADICLYGHLHRPAAWQVG 120

40 Query: 121 KTIFINPGSVLQPRGPINEKLYAVVTITDSKVLVEYYTRQHQPYPNLTKELSR 173  
+T+F+NPGSV QPRG INEKLVA V +TD+++ V+Y+TR H+ YP+L+KE R  
Sbjct: 121 QTLFMNPGSVTQPRGEINEKLYARVELTDTQIKVDYFTRDHKLYPSLSKEFKR 173

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1541

45 A DNA sequence (GBSx1632) was identified in *S.agalactiae* <SEQ ID 4749> which encodes the amino acid sequence <SEQ ID 4750>. This protein is predicted to be HAM1 family protein. Analysis of this protein sequence reveals the following:

50 Possible site: 23  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----  
bacterial cytoplasm --- Certainty=0.1218(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14796 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]

-1710-

Identities = 96/189 (50%), Positives = 130/189 (67%), Gaps = 1/189 (0%)

Query: 128 LIATHNEGKTKEFRELFGLKLVENLNDYPDLPEVEETGMTFEENARLKAEITISKLTGK 187  
 +IATHN GK KEF+E+ G V++L + E+EETG TFEENA +KAE ++K K  
 Sbjct: 8 IIATHNPGKVKEFEILEPRGYDVKSLAEIGFTTEEIEETGHTFEENAIMKAEAVAKAVNK 67

Query: 188 MVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDKERRSAQFHTTL 247  
 MVI+DDSG L +D LGG PGV+SAR++G D N K+L EL + +KE+R+A+F L  
 Sbjct: 68 MVIADDSGLSIDNLGGRPGVVSARYAGEQKDDQANIEKVLSELKGI-EKEQRTARFRCL 126

Query: 248 VVSAPNKESLVVEAEWPGYIGTEPKGENGFYDPLFIVGEGSRTAAELSAQEKNNLSHRG 307  
 VS P +E+ VE GYI EP+GE GFGYDP+FIV + +T AEL++ EKN +SHR  
 Sbjct: 127 AVSIPGEETKTVEGHVEGYIAEEPRGEYGFYDPIFIVKDKDKTMAELTSDEKKNKISHRA 186

Query: 308 QAVRKLMEV 316  
 A++KL ++  
 Sbjct: 187 DALKKLSKL 195

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4751> which encodes the amino acid  
 sequence <SEQ ID 4752>. Analysis of this protein sequence reveals the following:

Possible site: 39  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 214/325 (65%), Positives = 253/325 (77%), Gaps = 5/325 (1%)

Query: 1 MTKTIFESKTEGNWFLGSGFQAFNYFTCFG-NDESYEAIQDVFHRLSLTLKVE---GLQLH 56  
 M++ I+E K E NWF+G N + +G + + I D+ + +TL E G +  
 Sbjct: 14 MSEKIYKYDENNWFIGNKTHNLISGWGVKHTTIKKIDDLGIAATLDWENPKGYDVS 73

Query: 57 VVQMTSDFQLLAFLVDMINQEYSRHIKVTQHKGAILVSEDDQLFLVHLPKEGTSLEKFFD 116  
 VV+ S L+ F++DMINQE R IKVT H G IL+ E+ +L V+LP+ G S FF  
 Sbjct: 74 VVRHQSPSLITFIIDMINQETQREIKVTPHAGTILLMENAKLLAVYLPPEGVSTATFF- 132

Query: 117 LKNDNNFGDTILIIATHNEGKTKEFRELFGLKLVENLNDYPDLPEVEETGMTFEENARL 176  
 ++ FGD ILIAT NEGKTKEFR LFG+LG +VENLNDYP+LPEV ETG TFEENARL  
 Sbjct: 133 ATSEQGFGDIILIIATRNEGKTKEFRNLFGQLGYRVENLNDYPELPEVAETGTTFEENARL 192

Query: 177 KAETISKLTGKMVISDDSGLKVDALGGLPGVWSARFSGPDATDARNNAKLLHELAMVFDK 236  
 KAETIS+LTGKMV++DDSGLKVDALGGLPGVWSARFSGPDATDA+NNAKLLHELAMVFD+  
 Sbjct: 193 KAETISRLTGKMLADDSGLKVDALGGLPGVWSARFSGPDATDAKNNAKLLHELAMVFDQ 252

Query: 237 ERRSAQFHTTLVVSAPNKESLVVEAEWPGYIGTEPKGENGFYDPLFIVGEGSRTAAELS 296  
 ++RSAQFHTTLVV+APNK+SLVVEA+WPGYI T+PKGGENGFYD+FIVGE AAEL  
 Sbjct: 253 KKRSQFHTTLVVAAPNKDSLVEADWPGYIATQPKGENGFYDVFIVGETGHHAEELE 312

Query: 297 AQEKNNLSHRGQAVRKLMEVFPKWQ 321  
 A +KN LSHRGQAVRKLMEVFP WQ  
 Sbjct: 313 ADQKNQLSHRGQAVRKLMEVFPWQ 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1711-

**Example 1542**

A DNA sequence (GBSx1633) was identified in *S.agalactiae* <SEQ ID 4753> which encodes the amino acid sequence <SEQ ID 4754>. This protein is predicted to be glutamate racemase (murl). Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.86    Transmembrane  114 - 130 ( 114 - 130)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10067> which encodes amino acid sequence <SEQ ID 10068> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
  Identities = 160/267 (59%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

20  Query: 27  MDSRPIGFLDSGVGGLTVVKEMFRQLPEEEVIFIGDQARAPYGP RPAPQQIREFTWQM VNF 86
      M + IGF+DSGVGGLTVVKE RQLP E + ++GD AR PYGPRP Q+R+FTW+M +F
  Sbjct: 1  MKKQAIGFIDSGVGLTVVKEAMRQLPNESIYYVGD TARC PYGPRPEDQVRKFTWEMTHF 60

25  Query: 87  LLTKNVKMI VIACNTATATAVAWQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIIIGTPMT 146
      LL KN+KM+VIACNTATA A ++IK+KL IPV+GVILPG+ AAIK+T+ ++G+IGT T
  Sbjct: 61  LLDKNIKMLVIACNTATAAALKDIKKLAIPVIGVILPGSRAAIKATHTNRIGVIGTEGT 120

  Query: 147  VKSDAYRQKI QALS PNTAVVSLACPKFVPVIVESNQMSSSLAKKV VYETLSPLVGK-LDTL 205
      VKS+ Y++ I + V SLACPKFVP+VESN+ SS++AKKV ETL PL + LDTL
  Sbjct: 121  VKSNQYKKMIHSDTKALVTS LACPKFVPLVESNEYSSAI AKVVAETLRPLKNEGLDTL 180

30  Query: 206  ILGCTHYPLLRPIIQNV MGAEVKLIDSGAETVRDISVLLNYFEINHWNQKH-GGHHFYT 264
      ILGCTHYPLLRPIIQN +G V LIDSGAETV ++S +L+YF + + QNK +FYT
  Sbjct: 181  ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLA VDSQNKEKAERNFYT 240

35  Query: 265  TASP KGFKEIABQWLS-QEINVERIVL 290
      T S + F IA +WL ++ VE I L
  Sbjct: 241  TGSSQMFHAIASEWLQLDDLA VEHITL 267

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4755> which encodes the amino acid sequence <SEQ ID 4756>. Analysis of this protein sequence reveals the following:

```

45  Possible site: 19
   >>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -1.70    Transmembrane  88 - 104 ( 86 - 104)

50  ----- Final Results -----
      bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:AAF72713 GB:AF263927 glutamate racemase [Carnobacterium sp. St2]
  Identities = 149/267 (55%), Positives = 202/267 (74%), Gaps = 3/267 (1%)

55  Query: 1  MDTRPIGFLDSGVGGLTVVCELIRQLPHEKIVYIGDSARAPYGP RP KQIKEYTWELVNF 60
      M + IGF+DSGVGGLTVV E +RQLP+E I Y+GD+AR PYGPRP+ Q++++TWE+ +F
  Sbjct: 1  MKKQAIGFIDSGVGLTVVKEAMRQLPNESIYYVGD TARC PYGPRPEDQVRKFTWEMTHF 60

  Query: 61  LLTQNVKMI VFACNTATATAVAWEEVKAALDIPVLGVVLP GASAAIKSTTKGQVGVIGTPMT 120

```

-1712-

LL +N+KM+V ACNTATA A +++K L IPV+GV+LPG+ AAIK+T ++GVIGT T  
 Sbjct: 61 LLDKNIKMLVIACNTATAAALKDIKKLAIPVIGVILPGSRAAIKATHINRIGVIGTEGT 120

Query: 121 VASDIYRKIKIQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYDSLAPLVGK-IDTL 179  
 V S+ Y+K I V SLACPKFVP+VESNE S+IAKK+V ++L PL + +DTL

Sbjct: 121 VKSNQYKKMIHSKDTKALVTSIACPKFVPLVESNEYSSAIKKVVAETLRPLKNEGLDTL 180

Query: 180 VLGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDIN-GNYHQKAVEHRFFT 238  
 +LGCTHYPLLRPIIQN +G SV LIDSGAE V ++S +L+YF++ + +++ E F+T

Sbjct: 181 ILGCTHYPLLRPIIQNTLGDSVTLIDSGAETVSEVSTILDYFNLAVDSONKEKAERNFYT 240

Query: 239 TANPEIFQEIASIWLK-QKINVEHVTL 264  
 T + ++F IAS WL+ + VEH+TL

Sbjct: 241 TGSSQMFHAIASEWLQLDDLAHEHITL 267

An alignment of the GAS and GBS proteins is shown below.

Identities = 195/264 (73%), Positives = 231/264 (86%)

Query: 27 MDSRPIGFLDSGVGGLTVVKEMFRQLPEEEVIFIGDQARAPYGPRPAQQIREFTWQMVNF 86  
 MD+RPIGFLDSGVGGLTVV E+ RQLP E++++IGD ARAPYGPRP +QI+E+TW++VNF

Sbjct: 1 MDRPIGFLDSGVGGLTVVCELIRQLPHEKIVYIGDSARAPYGPRPKQIKEYTWELVNF 60

Query: 87 LLTKNVKMIVACNTATAVAWQEIKEKLDIPVLGVILPGASAAIKSTNLGKVGIGTPMT 146  
 LLT+NVKMIV ACNTATAVAW+E+K LDIPVLGV+LPGASAAIKST G+VG+IGTPMT

Sbjct: 61 LLTQNVKMIVFACNTATAVAWEEVKAALDIPVLGVVLPASAAIKSTTKGQVGVIGTPMT 120

Query: 147 VKSDAYRQKIQALSPTAVVSLACPKFVPVIVESNQSSSLAKKVYETLSPLVGKLDTLI 206  
 V SD YR+KIQ L+P+ V SLACPKFVPVIVESN+M SS+AKK+VY++L+PLVGK+DTL+

Sbjct: 121 VASDIYRKIKIQLLAPSIQVRSIACPKFVPVIVESNEMCSSIAKKIVYDSLAPLVGKIDTLV 180

Query: 207 LGCTHYPLLRPIIQNVMGAEVKLIDSGAETVRDISVLLNYFEINHNWQNKHGGHHFYTTA 266  
 LGCTHYPLLRPIIQNVMG VKLIDSGAE VRDISVLLNYF+IN N+ K H F+TTA

Sbjct: 181 LGCTHYPLLRPIIQNVMGPSVKLIDSGAECVRDISVLLNYFDINGNYHQKAVEHRFFTTA 240

Query: 267 SPKGFKEIAEQWLSQEINVERIVL 290  
 +P+ F+EIA WL Q+INVE + L

Sbjct: 241 NPEIFQEIASIWLKQKINVEHVTL 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1543

A DNA sequence (GBSx1634) was identified in *S. agalactiae* <SEQ ID 4757> which encodes the amino acid sequence <SEQ ID 4758>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.36 Transmembrane 3 - 19 ( 1 - 27)

----- Final Results -----

bacterial membrane --- Certainty=0.5543(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13675 GB:Z99113 alternate gene name: yoxG [Bacillus subtilis]  
 Identities = 26/72 (36%), Positives = 42/72 (58%)

Query: 1 MSITIWILLIIVALFGGLVGGIFIAKQIEKEIGEHPRLTPDAIREMMSQMGQKPSAKV 60  
 M++ + IL+ +VAL G+ G FIARK + + ++P + +R MM QMG KPS+ K+

Sbjct: 1 MTLWVGILVGVVALLIGVALGFFIARKYMMSYLKKNPPIEQMLRMMMMQMGKPSQKKI 60

Query: 61 QQTYRNIVKHAK 72

-1713-

Q + + K  
 Sbjct: 61 NQMMKAMNNQTK 72

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4759> which encodes the amino acid sequence <SEQ ID 4760>. Analysis of this protein sequence reveals the following:

Possible site: 15  
 >>> Seems to have an uncleavable N-term signal seq  
 INTEGRAL Likelihood = -10.72 Transmembrane 7 - 23 ( 1 - 27)

----- Final Results -----  
 bacterial membrane --- Certainty=0.5288(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 62/79 (78%), Positives = 69/79 (86%)

Query: 1 MSITIWILLIIVALFGGLVGGIFIARKQIEKEIGEHPRLTPDAIREMMSQMGQKPSEAKV 60  
 MS IWILL+IVAL G+ GGIFIARKQIEKEIGEHPRLTP+AIEMMSQMGQKPSEAK+  
 Sbjct: 1 MSTAIWILLIIVALGVGVFGGIFIARKQIEKEIGEHPRLTPEAIREMMSQMGQKPSEAKI 60  
 Query: 61 QQTYRNIVKHAKTAIKTKK 79  
 QQTYRNI+K +K A+ K  
 Sbjct: 61 QQTYRNIIKQSKAAVSKGK 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1544

A DNA sequence (GBSx1635) was identified in *S.agalactiae* <SEQ ID 4761> which encodes the amino acid sequence <SEQ ID 4762>. Analysis of this protein sequence reveals the following:

Possible site: 57  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.86 Transmembrane 82 - 98 ( 79 - 103)

----- Final Results -----  
 bacterial membrane --- Certainty=0.4142(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1545

A DNA sequence (GBSx1636) was identified in *S.agalactiae* <SEQ ID 4763> which encodes the amino acid sequence <SEQ ID 4764>. Analysis of this protein sequence reveals the following:

Possible site: 30  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -10.83 Transmembrane 56 - 72 ( 50 - 105)  
 INTEGRAL Likelihood = -7.27 Transmembrane 27 - 43 ( 17 - 48)  
 INTEGRAL Likelihood = -6.26 Transmembrane 76 - 92 ( 73 - 105)  
 INTEGRAL Likelihood = -4.83 Transmembrane 119 - 135 ( 118 - 141)  
 INTEGRAL Likelihood = -1.65 Transmembrane 160 - 176 ( 160 - 176)

-1714-

## ----- Final Results -----

5 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8837> which encodes amino acid sequence <SEQ ID 8838> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4765> which encodes the amino acid sequence <SEQ ID 4766>. Analysis of this protein sequence reveals the following:

Possible site: 19

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

15 INTEGRAL Likelihood = -10.99 Transmembrane 45 - 61 ( 37 - 94)  
 INTEGRAL Likelihood = -7.06 Transmembrane 74 - 90 ( 62 - 94)  
 INTEGRAL Likelihood = -3.45 Transmembrane 110 - 126 ( 108 - 130)  
 INTEGRAL Likelihood = -2.18 Transmembrane 149 - 165 ( 149 - 165)  
 INTEGRAL Likelihood = -1.91 Transmembrane 21 - 37 ( 20 - 37)

20 ----- Final Results -----

bacterial membrane --- Certainty=0.5394(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

Identities = 111/173 (64%), Positives = 145/173 (83%)

30 Query: 12 MSKKTQMVSYTSILVAFAIMIPIIMPAKIIIGPASFTLASHVPLFLSIFISVPVAILVA 71  
 M+KK TQ++YTSILVAFAI+IPIIMP K+IIGPASFTLASHVPLFL+IF+S+PVAILVA  
 Sbjct: 1 MTKKPTQLIAYTSILVAFAILIPIIMPLKLIIGPASFTLASHVPLFLAIFMSIPVAILVA 60

35 Query: 72 LGTGLGFLLAGFPPIVIVLRALSHIGFALIAAFLIKSKPSLLMSKQTLFVAVAINIIHGL 131  
 LGT LGFLLAG P++IVLRALSH+ FA++AA+ + KP L+ S + FA IN+IHGL  
 Sbjct: 61 LGTTLGFLLAGLPLIIVLRALSHLLFAILAAWLSRKQLMTSAVKCFSFAFFINVIHGL 120

40 Query: 132 LEFTIVYIITMTSNSSSTYLWSLFLIGLGSLLHGLVDFYIALFIWKWMTQKL 184  
 EF+ VYI+T T+ +S +Y WS+ LIGLGSLLHG++DFY+AL +W+++ + L  
 Sbjct: 121 AEFLVVYILTATTATSMYSFWSMLGLIGLGSLLHGLIDFYALVLWRFLAKNL 173

A related GBS gene <SEQ ID 10789> and protein <SEQ ID 10790> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3

SRCFLG: 0

45 McG: Length of UR: 24  
 Peak Value of UR: 3.16  
 Net Charge of CR: 2

McG: Discrim Score: 12.56

GvH: Signal Score (-7.5): -0.16

50 Possible site: 19

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

Amino Acid Composition: calculated from 20

ALOM program count: 5 value: -10.83 threshold: 0.0

55 INTEGRAL Likelihood = -10.83 Transmembrane 45 - 61 ( 39 - 94)  
 INTEGRAL Likelihood = -6.26 Transmembrane 65 - 81 ( 62 - 94)  
 INTEGRAL Likelihood = -4.83 Transmembrane 108 - 124 ( 107 - 130)  
 INTEGRAL Likelihood = -1.65 Transmembrane 149 - 165 ( 149 - 165)  
 INTEGRAL Likelihood = -0.27 Transmembrane 24 - 40 ( 24 - 40)  
 PERIPHERAL Likelihood = 0.42 86

-1715-

modified ALOM score: 2.67  
icml HYPID: 7 CFP: 0.533

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5331(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1546

A DNA sequence (GBSx1637) was identified in *S.agalactiae* <SEQ ID 4767> which encodes the amino acid sequence <SEQ ID 4768>. This protein is predicted to be transcriptional regulator, biotin repressor family. Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2237(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14749 GB:Z99118 yrxA [Bacillus subtilis]

Identities = 72/165 (43%), Positives = 112/165 (67%), Gaps = 2/165 (1%)

Query: 6 RRENILTTTLKGTKEAISASTLAKIFSVSRQVIVGDIALLRQQCDIISTPKGYL-MSSAL 64

RR+ +L LK +K ++ LAK +VSRQVIV DI+LL+A+ II+T +GY+ M +A

Sbjct: 12 RRDQLLLWLKESKSPITGGELAKKANVSRQVIVQDISLLKAKNPPIATSQGVYMDAA 71

Query: 65 STHQFTARLV-CQHIEQTEEELEIILRYQGIIMNVEVEHPYIGMLTAPLNIQSQKDIDN 123

HQ R++ C HG E+TEEEEL++I+ + +V++EHP+YG LTA + + ++K++ +

Sbjct: 72 QQHQQAERIIACLHGPERTEEEELQLIVDEGVTVKDVKIEHPVYGDLTAAIQVGTKEVSH 131

Query: 124 FTAKLKVSNAELLSSLTDGLHMHMISCQDQSVFDQICEALKKAGI 168

F K+ +NA LS LTDG+H H ++ D+ DQ C+AL++AGI

Sbjct: 132 FIKKINSTNAAYLSQLTDGVHLHTLTAPDEHRIDQACQALEEAGI 176

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4769> which encodes the amino acid sequence <SEQ ID 4770>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2971(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 109/170 (64%), Positives = 136/170 (79%)

Query: 1 MKAQERRENILTTTLKGTKEAISASTLAKIFSVSRQVIVGDIALLRQQCDIISTPKGYLM 60

MKA++RR+ I+ L ++A+SA+ L K+ VSRQVIVGDIALLRQQ DIISTPKGY+M

Sbjct: 1 MKAEDRRQKIIIECLNSEQKAVSATRLGKLLGVSRQVIVGDIALLRQQIDIISTPKGYIM 60

Query: 61 SSALSTHQFTARLV-CQHIEQTEEELEIILRYQGIIMNVEVEHPYIGMLTAPLNIQSQKD 120

-1716-

S+AL +HQF AR+VCQH +E+T++ELEIIL +QGII VEVEHPYGM+TAPLNI++ D  
 Sbjct: 61 STALYSHQFCARIVCQHNVETKKELEIILAHQGIITTVVEVEHPYGMITAPLNIKTHSD 120

Query: 121 IDNFTAKLKVSNAELLSSSLTDGLHMHMISCQDQSVFDQICEALKKAGILY 170  
 + NF +KL S AELLSSSLT+GLH+H+ISC Q F I L+ AGILY  
 Sbjct: 121 VTNFMSKLSQSKAELLSSSLTEGLHSHLISCPSQEAFLAIKHDLELAGILY 170

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

## 10 Example 1547

A DNA sequence (GBSx1638) was identified in *S.agalactiae* <SEQ ID 4771> which encodes the amino acid sequence <SEQ ID 4772>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have no N-terminal signal sequence

15	INTEGRAL	Likelihood = -8.44	Transmembrane	143 - 159 ( 138 - 165)
	INTEGRAL	Likelihood = -8.17	Transmembrane	164 - 180 ( 160 - 184)
	INTEGRAL	Likelihood = -7.17	Transmembrane	56 - 72 ( 53 - 78)
	INTEGRAL	Likelihood = -5.63	Transmembrane	24 - 40 ( 21 - 44)
	INTEGRAL	Likelihood = -4.94	Transmembrane	113 - 129 ( 108 - 131)
20	INTEGRAL	Likelihood = -2.39	Transmembrane	86 - 102 ( 86 - 103)
	INTEGRAL	Likelihood = -1.06	Transmembrane	203 - 219 ( 203 - 219)

----- Final Results -----

25	bacterial membrane	--- Certainty=0.4376(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10069> which encodes amino acid sequence <SEQ ID 10070> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC18360 GB:AF064763 putative membrane spanning protein

[Lactococcus lactis subsp. cremoris]

Identities = 97/188 (51%), Positives = 133/188 (70%)

35 Query: 38 IMLYMFPQNMIAIMQKMPGLYFGAIIELVLVVFASGAARRNTPAALPLFLIYSALNGFT 97  
 IM+ F NM AI+Q I+ LV+V G A +N+ ALP+F+ Y+A GF  
 Sbjct: 1 IMITFFLDNMRAILQSGSLFLLVLWIIPLVMVVSLLQGLAMKNSKMALPIFIGYAAFMGFL 60

40 Query: 98 LSFIIARYTQTTVLQAFITSAAVFFAMALIGAKTKKDLSGMRKALMAALIGILIASLVNL 157  
 +SF + YT T + AFIT++A+FF +++ G TK++LSGM KAL A+ G+++A L+NL  
 Sbjct: 61 ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTKRNLSGMGKALGVAVWGLIVAMLLNL 120

Query: 158 FIGSGGMSYIISIVCVIIFSGLIAVDNQMIKYVYNSQGGQVADGWAVSMALSLYLDFFINL 217  
 F S G++ +IS+V V+IFSGLIA+DNQ I VYN+ GQV+DGWA+SMALSLYLDFFIN+  
 45 Sbjct: 121 FFASTGLTILISLVGVVIFSGLIAVDNQKITQVYNAHNGQVSDGWAISMALSLYLDFFINM 180

Query: 218 FLNILRLF 225  
 FL +LRLF  
 Sbjct: 181 FLFLRLRF 188

50

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4773> which encodes the amino acid sequence <SEQ ID 4774>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

55	INTEGRAL	Likelihood = -8.97	Transmembrane	143 - 159 ( 138 - 165)
	INTEGRAL	Likelihood = -5.89	Transmembrane	164 - 180 ( 160 - 184)
	INTEGRAL	Likelihood = -5.68	Transmembrane	56 - 72 ( 55 - 77)
	INTEGRAL	Likelihood = -4.78	Transmembrane	113 - 129 ( 110 - 130)

-1717-

INTEGRAL Likelihood = -2.81 Transmembrane 203 - 219 ( 203 - 222)  
 INTEGRAL Likelihood = -2.76 Transmembrane 24 - 40 ( 23 - 41)  
 INTEGRAL Likelihood = -2.76 Transmembrane 86 - 102 ( 86 - 104)

5 ----- Final Results -----  
           bacterial membrane --- Certainty=0.4588(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:AAC18360 GB:AF064763 putative membrane spanning protein  
 [Lactococcus lactis subsp. cremoris]  
 Identities = 90/189 (47%), Positives = 133/189 (69%)

15 Query: 38 LMLYPFRENLSILVNQPMIYYGAIIELILVVFVASSAARKNTPAALPIFLIYSALNGFT 97  
           +M+ F +N+ +IL + + II L++V A KN+ ALPIF+ Y+A GF  
       Sbjct: 1 IMITFFLDNMRILQSGSLFLLVLWIIPLVMVVSLOGLAMKNSKMALPIFIGYAAFNGFL 60

20 Query: 98 LSFIIVAYAQTTFVQAFLLSSAAVFFFAMSIIGVTKRDMGSLRKAMFAALIGVVVASLINL 157  
           +SF ++ Y T + AF++++A+FF +S+ G TKR++SG+ KA+ A+ G++VA L+NL  
       Sbjct: 61 ISFTLLMYTATDITLAFITASAMFFGLSVYGRFTKRNLGSGMGKALGVAVWGLIVAMLLNL 120

25 Query: 158 FIGSGMMSYVISVISVLIFSGLIASDNQMIKRVYQATNGQVGDGWAAMALSLEYLDFINL 217  
           F S ++ +IS++ V+IFSGLIA DNQ I +VY A NGQV DGWA++MALSLYLDFFIN+  
       Sbjct: 121 FFASTGLTILISLVGVVIFSGLIADWNQKITQVYNHNGQVSDGWAISMALSLYLDFFINM 180

30 Query: 218 FISLLRIFG 226  
           F+ LLR+FG  
       Sbjct: 181 FLFLRLRIFG 189

An alignment of the GAS and GBS proteins is shown below.

Identities = 167/229 (72%), Positives = 202/229 (87%)

35 Query: 1 MNDNVIYTQSDSGLNQFFAKIYGLVGIGVGLSAAVSAIMLYMFPQNMIAIMQKMPGLYFG 60  
           MND+VIYTQSD GLNQFFAKIY LVG+GVGLSA VS +MLY F +N+I+I+ P +Y+G  
       Sbjct: 1 MNDHVIYTQSDVGLNQFFAKIYSLVGMGVGLSAFVSYLMLYPFRENLSILVNQPMIYYG 60

40 Query: 61 AIIELVLVVFVASSGAARRNTPAALPLFLIYSALNGFTLSFIIARYTQTTVLQAFITSAAV 120  
           A I+EL+LVFVAS AAR+NTPAALP+FLIYSALNGFTLSFII Y QTTV QAF++SAAV  
       Sbjct: 61 AAIIEILILVVFVASSAARKNTPAALPIFLIYSALNGFTLSFIIVAYAQTTFVQAFLLSSAAV 120

45 Query: 121 FFAMALIGAKTKKDLGSGMRKALMAALIGILIASLVNLFISGGMSYIISIVCVIIFSGLI 180  
           FFAM++IG KTK+D+SG+RKA+ AALIG+++ASL+NLFIGSG MSY+IS++ V+IFSGLI  
       Sbjct: 121 FFAMSIIGVTKRDMGSLRKAMFAALIGVVVASLINLFISGGMMSYVISVISVLIFSGLI 180

50 Query: 181 AYDNQMIKYVYNSQGGQVADGWAVSMALSLYLDFFINLFLNLRFLARND 229  
           A DNQMIK VY + GQV DGWAV+MALSLYLDFFINLF+++LR+F RND  
       Sbjct: 181 ASDNQMIKRVYQATNGQVGDGWAAMALSLYLDFFINLFISLLRIFGRND 229

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1548

A DNA sequence (GBSx1639) was identified in *S.agalactiae* <SEQ ID 4775> which encodes the amino acid sequence <SEQ ID 4776>. Analysis of this protein sequence reveals the following:

55 Possible site: 44  
       >>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----  
           bacterial cytoplasm --- Certainty=0.2495(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1718-

A related GBS nucleic acid sequence <SEQ ID 10071> which encodes amino acid sequence <SEQ ID 10072> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4777> which encodes the amino acid sequence <SEQ ID 4778>. Analysis of this protein sequence reveals the following:

Possible site: 60  
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----  
bacterial cytoplasm --- Certainty=0.3277(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 15 An alignment of the GAS and GBS proteins is shown below.

Identities = 127/163 (77%), Positives = 141/163 (85%)

Query: 7 YQDDKDFMDLVGHLIDHPRFQKLEAIVQHHHSTRLEHSINVSYTSYKIAKKFGWDASSTA 66  
Y +DK++M+ VGH LI HPRFQKL IVQH HSTRLEHSINVSY+SYK+AK+FGWDA STA  
20 Sbjct: 3 YTEDKEYMEHVGH LIAHPRFQKLSHIVQHSHSTRLEHSINVSYSSYKLAKRFGWDAKSTA 62  
Query: 67 RGGLLHDDFFYYDWRVTKFNKSHAWVHPRIAVRNARKLTDLNAREEDIILKHMWGATIAPP 126  
RGGLLHDDFFYYDWRVTKFNK HAWVHPRIAVRNA+KLT+LN +EEDIILKHMWGATIA P  
Sbjct: 63 RGGLLHDDFFYYDWRVTKFNGHAWVHPRIAVRNAKLT+LN+KKEEDIILKHMWGATIAPP 122  
25 Query: 127 RYKESYIVTMVDKYWAVREASRPLKRIFFKPIRFSRKFLGSHN 169  
RYKESYIVTMVDKYWAV+EA PL++ + RK L SHN  
Sbjct: 123 RYKESYIVTMVDKYWAVKEAVTPLRQKWSNRRFLRRKTLQSHN 165

- 30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1549

A DNA sequence (GBSx1640) was identified in *S.agalactiae* <SEQ ID 4779> which encodes the amino acid sequence <SEQ ID 4780>. Analysis of this protein sequence reveals the following:

35 possible site: 37  
>>> Seems to have no N-terminal signal sequence  
INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 ( 212 - 229)

40 ----- Final Results -----  
bacterial membrane --- Certainty=0.2211(Affirmative) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 A related GBS nucleic acid sequence <SEQ ID 9413> which encodes amino acid sequence <SEQ ID 9414> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14825 GB:Z99118 similar to rRNA methylase [Bacillus subtilis]  
Identities = 96/228 (42%), Positives = 143/228 (62%), Gaps = 5/228 (2%)

50 Query: 3 QKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVT-ETAIDR-LRKPERAIVVTDVVKELT 60  
+++ + +++LIEG HL EEA K I V ET I L + +++D +T  
Sbjct: 22 KERTKTNTFLIEGEHLVEEALKSPGIVKEILVKDETRIPSDLETGIQCYMLSEDAFSAVT 81  
Query: 61 DSQTPQGIVAEIAFQETRWTDIKKGRFLVLEDDVQDPGNLGTMVRTADAANFDAVFLSQKS 120

-1719-

+++TPQ I A E + +K L+++ VQDPGNLGT+RTADAA DAV L +  
 Sbjct: 82 ETETPQQIAAVCHMPPEEKLATARK--VLLIDAVQDPGNLGT+MIRTADAAGLDAVVLGDGT 139

Query: 121 ADLYNQKTLRSMQGSHPVFRVEIEQFVNFC+KAEGITMIATTLSEQSVNYKNLPKYDY 180  
 AD +N KTLRS QGSHPV R + +V+ KAEG+ + T L + Y+ +P+ +  
 Sbjct: 140 ADAFNGKTLRSAQGSHPVFRVRRNLPSYVDELKAEGVKVYGTAL-QNGAPYQEIPQSES 198

Query: 181 FALIMGNEGQGISKTMTEADVLAHIEMPGQAESLNVAAGVVIFSL 228  
 FALI+GNEG G+ + E+ D+ ++ + GQAESLNVAAG +++++ L  
 Sbjct: 199 FALIVGNEGAGVDAALLEKTDNLNLYVPLYGQAESLNVAAGAILVYHL 246

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4781> which encodes the amino acid sequence <SEQ ID 4782>. Analysis of this protein sequence reveals the following:

Possible site: 61  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -2.97 Transmembrane 229 - 245 ( 228 - 245)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2190(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 141/229 (61%), Positives = 178/229 (77%)

Query: 1 MLQKKYRKSSYLIEGWHLFEEAEKYGAQFLNIFVTETAIDRLRKPERAIVVTDVVKELT 60  
 +LQKK+RK SYLIEGWHLFEEA+K G F +IFV E ++RL + ++V+ VLKELT  
 Sbjct: 17 LLQKKHRKQSYLIEGWHLFEEAQSGQVFRHIFVLEEMVERLAGEQELVIVSPQVLKELT 76

Query: 61 DSQTPQGIVAEIAFQETRTWDIKGRFLVLEDVQDPGNLGT+RTADAAANFVFLSOKS 120  
 DS +PQGIVAE+ + + KG++LVLEDVQDPGNLGT++RTADAA FD VFLS+KS  
 Sbjct: 77 DSPSPQGIVAEVEIPKLAFFPSDYKGKYLLEDVQDPGNLGTIIRTADAARFDGVFLSEKS 136

Query: 121 ADLYNQKTLRSMQGSHPVFRVEIEQFVNFC+KAEGITMIATTLSEQSVNYKNLPKYDY 180  
 AD+YNQKTLRSMQGSHPVFR++R ++ Q + ++ATTL++SV+YK+L ++  
 Sbjct: 137 ADIYNQKTLRSMQGSHPVFRVTDVYQLCRELQYETPILATTLKKSVVDYKSLTHHER 196

Query: 181 FALIMGNEGQGISKTMTEADVLAHIEMPGQAESLNVAAGVVIFSLI 229  
 AL++GNEGQGIS M AD L HI MPGQAESLNVAAG++IFSLI  
 Sbjct: 197 LALVLGNEGQGISAEAAALADQLVHITMPGQAESLNVAAGILIFSLI 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8839> and protein <SEQ ID 8840> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 7  
 McG: Discrim Score: -7.98  
 GvH: Signal Score (-7.5): -3.86  
 Possible site: 37  
 >>> Seems to have no N-terminal signal sequence  
 ALOM program count: 1 value: -3.03 threshold: 0.0  
 INTEGRAL Likelihood = -3.03 Transmembrane 213 - 229 ( 212 - 229)  
 PERIPHERAL Likelihood = 5.14 149  
 modified ALOM score: 1.11

\*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

```

ORF02468(259 - 984 of 1287)
EGAD|107730|BS2859(4 - 246 of 248) hypothetical protein {Bacillus subtilis}
GP|1770029|emb|CAA99602.1||Z75208 hypothetical protein {Bacillus subtilis}
GP|2635330|emb|CAB14825.1||Z99118 similar to rRNA methylase {Bacillus subtilis}
PIR|G69984|G69984 rRNA methylase homolog ysgA - Bacillus subtilis
%Match = 20.3
%Identity = 43.0 %Similarity = 62.3
Matches = 105 Mismatches = 87 Conservative Sub.s = 47

```

186 216 246 276 306 330 360 390  
A\*RNPTP\*TRPETIK\*TFFIT\*PLF\*YNRKMTITIITSKSNLIKTKKLLQKKYR--KSSYLIEGWHLFEEAEKYGAQFL  
| | | : | | | | | : : | | | | | | |  
MKQIESAKNQKVKDWKKLHTKKERTKTNTFLIEGHLVEEALKSPGIVK  
10 20 30 40

417            444            474            504            534            564            594            624  
 NIFVT-ETAI-DRLRKPERAIVVTDDVLKELTDSQT PQGIVAEIAFQETRWTDIKKGRFLVLEDVQDPGNLGTMVRTADA  
 |:| |:| |:| |:| : ::::| : ::::|:| |:| |:| : :|::| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|  
 EILVKDETRIPSDLETGIQCYMLSEDAFSAVTTETETPQQTAAVCHMPEEKLA--TARKVLLIDAVQDPGNLGTMTIRTADA  
 60            70            80            90            100            110            120

[illegible]

894 924 954 984 1014 1044 1074 1104  
GQGISKMTTEEADVLAHTEMPPGQAESLNVAVAGVVFSLI\*VHML\*YPQRGDYNEKVSRR\*GLHGFRSPY\*PSTFPKT  
| |: : |: : :: : ||||| ||||| : :: :  
GAGVDAALLEKTDNLNLYVPLYGQAESLNVAVAAAILVYHLRG  
220 230 240

SEQ ID 8840 (GBS430) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 5; MW 29kDa).

GBS430-GST was purified as shown in Figure 220, lane 8.

### Example 1550

A DNA sequence (GBSx1641) was identified in *S.agalactiae* <SEQ ID 4783> which encodes the amino acid sequence <SEQ ID 4784>. This protein is predicted to be acylphosphatase (acyP). Analysis of this protein sequence reveals the following:

Possible site: 48  
>>> Seems to have an uncleavable N-term signal seq

```

----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside  --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10073> which encodes amino acid sequence <SEQ ID 10074> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD36630 GB:AE001801 acylphosphatase, putative [Thermotoga maritima]  
Identities = 35/88 (39%), Positives = 52/88 (58%), Gaps = 3/88 (3%)

Query: 24 MKKVHLIVSGRVQGVGFRYATYSLALEIGDIYGRVWNDDGTVEILAQSTDSNKMTQFIQ 83  
MK + + V G VQGVGFRY T +A +G + G V N DDG+V I A+ D N + +F+

-1721-

Sbjct: 1 MKALKIRVEGIVQGVGFRYFTRRVAKSLG-VKGYVMNDDGSVFIHAEG-DENALRRFLN 58

Query: 84 KIRKGPSKWSKVITYVDIKLDFDFDFND 111  
 ++ KGP + VT V ++ + + DF

5 Sbjct: 59 EVAKGPPA-AVVTNVSVEETTPGYEDF 85

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4785> which encodes the amino acid sequence <SEQ ID 4786>. Analysis of this protein sequence reveals the following:

Possible site: 34  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2433(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 69/95 (72%), Positives = 85/95 (88%)

Query: 19 KRGQVMKKVHLIVSGRVQGVGFRYATYSLALEIGDIYGRVWNNDDGTVEILAQSTDSNKM 78  
 K +M+KV LIVSGRVQGVGFRYAT++LAL+IGDIYGRVWNN+DGTVEILAQS DS+K+

Sbjct: 7 KEALLMQKVRILIVSGRVQGVGFRYATHTLALDIGDIYGRVWNNDDGTVEILAQSKDSKI 66

Query: 79 TQFIQKIRKGPSKWSKVITYVDIKLDFDFDFNDFKM 113  
 FIQ++RKGPSKW+KVITYVD+ + NF+DF DF++

25 Sbjct: 67 ATFIQEVVRKGPSKWAKVITYVDVTMANFEDFQDFQI 101

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1551

A DNA sequence (GBSx1642) was identified in *S.agalactiae* <SEQ ID 4787> which encodes the amino acid sequence <SEQ ID 4788>. This protein is predicted to be membrane protein homolog (yidC). Analysis of this protein sequence reveals the following:

Possible site: 16  
 >>> May be a lipoprotein

INTEGRAL	Likelihood = -12.52	Transmembrane	60 - 76 ( 54 - 83)
INTEGRAL	Likelihood = -3.66	Transmembrane	178 - 194 ( 177 - 196)
INTEGRAL	Likelihood = -2.76	Transmembrane	140 - 156 ( 137 - 157)
INTEGRAL	Likelihood = -2.60	Transmembrane	216 - 232 ( 213 - 232)

----- Final Results -----  
 bacterial membrane --- Certainty=0.6010(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10075> which encodes amino acid sequence <SEQ ID 10076> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria  
 monocytogenes]

Identities = 82/222 (36%), Positives = 133/222 (58%), Gaps = 4/222 (1%)

Query: 44 PMANLITYFAQHQLGFGVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAFYKPLFPEPI 103  
 P + I + A+ G +G+AIII T+++R +I+PL L + KMA KP + I

55 Sbjct: 3 PFTSFIMFVAKFVGGNYGIAIIITLLIRALIMPLNLRTAKAQMGMSKMAVAKPEIDEI 62

-1722-

Query: 104 NERLRNAKTQEEKLAAQTELMTAQRENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGV 163  
 RL+ A ++EE+ Q E+M + ++ +GCLPLLIQMP A ++A R + +  
 Sbjet: 63 QARLKRATSKKEEQATIQKEMMAVYSKYNNINPMQ-MGCLPLLIQMPILMAFYAIRGSSEI 121

5 Query: 164 SSATFLGLNLGQKSLTLTVIIAILYFVQSWLSMQGVPDEQRQMKMTMYLMPIMMVFSI 223  
 +S TFL NLG + L +I ++Y Q ++SM G EQ++QMK + + PIM++F+S  
 Sbjet: 122 ASHTFLWFLNLGSPDMVLAIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMILFVSF 181

10 Query: 224 SLPASVALYWFIGGIFSIIQQLVT--TYVLK-PKLRKVEEE 262  
 + P+++ALYW +GG+F Q L+T Y+ K P+++ +EE  
 Sbjet: 182 TAPSALALYWAVGGLFLAGQTLTKKLYMNKHPEIKVMEQEE 223

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4789> which encodes the amino acid sequence <SEQ ID 4790>. Analysis of this protein sequence reveals the following:

15 Possible site: 31

>>> May be a lipoprotein

INTEGRAL	Likelihood = -9.55	Transmembrane	62 - 78 ( 54 - 82)
INTEGRAL	Likelihood = -2.81	Transmembrane	178 - 194 ( 177 - 195)
INTEGRAL	Likelihood = -0.90	Transmembrane	216 - 232 ( 215 - 232)

20

----- Final Results -----

bacterial membrane ---	Certainty=0.4821(Affirmative) < succ>
bacterial outside ---	Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm ---	Certainty=0.0000(Not Clear) < succ>

25

The protein has homology with the following sequences in the databases:

>GP:AAF03934 GB:AF139908 membrane protein homolog [Listeria monocytogenes]  
 Identities = 89/218 (40%), Positives = 132/218 (59%), Gaps = 2/218 (0%)

30 Query: 43 KPMSYFIDYFANNAGLYGLAIIIVTIIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEP 102  
 +P + FI + A G YG+AIII T+++R LI+PL L + KMA KP +  
 Sbjet: 2 QPPTSFIMFVAKFVGGNYGIAIIITTLIRALIMPLNLRATAQMGMSKMAVAKPEIDE 61

35 Query: 103 INKRIKQANSQEEKMAAQTELMAAQRAHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKG 162  
 I R+K+A S+EE+ Q E+MA + INP+ +GCLPLLIQMP A Y+A + +  
 Sbjet: 62 IQARLKRATSKKEEQATIQKEMMAVYSKYNNINPMQ-MGCLPLLIQMPILMAFYAIRGSSE 120

40 Query: 163 VSTSTFMGIDLGSRLVLTATIAALYFFQSWLSMMAVSEEQREQMKMTMYTMPIMMIFMS 222  
 +++ TF+ +LGS +VL I +Y Q ++SM+ S EQ++QMK + PIM++F+S  
 Sbjet: 121 IASHTFLWFLNLGSPDMVLAIAGLVYLAQYFVSMIGYSPEQKKQMKIIGLMSPIMILFVS 180

Query: 223 FSLPAGVGLYWLVGFFSIIQQLITTYYLLKPLHKQIK 260  
 F+ P+ + LYW VGG F Q L+T L + H +IK  
 45 Sbjet: 181 FTAPSALALYWAVGGLFLAGQTLTKKLYMNK-HPEIK 217

An alignment of the GAS and GBS proteins is shown below.

Identities = 203/309 (65%), Positives = 254/309 (81%), Gaps = 2/309 (0%)

50 Query: 1 MKKTLKRILFSSLSLSMLLLLLTGCVSDKAGKPYGVIWNTLGVPANLITYFAHQGLGF 60  
 +K TL RILFS L+LS+LL LTGCV D G P G+IW LG PM+ I YFA + GLG+  
 Sbjet: 1 LKLTNLRILFSGLALSILLTLTGCVGRDAHGNPKGMIWEFLGKPMYFIDYFANNAGLY 60

55 Query: 61 GVAIIIVTVIVRVVILPLGLYQSWKASYQAEKMAFYKPLFEPINERLRNAKTQEEKLAAQ 120  
 G+AIIIVT+IVR +ILPLGLYQSWKASYQ+EKMA+ KP+FEPIN+R++ A +QEEK+AAQ  
 Sbjet: 61 GLAIIIVTIIIVRTLILPLGLYQSWKASYQSEKMAFLKPVFEPINKRIKQANSQEEKMAAQ 120

Query: 121 TELMTAQRENGLSMFGGIGCLPLLIQMPFFSAIFFAARYTPGVSSATFLGLNLGQKSLTL 180  
 TELM AQR +G++ GGIGCLPLLIQMPFFSA++FAA+YT GVS++TF+G++LG +SL L  
 60 Sbjet: 121 TELMAAQRHGINPLGGIGCLPLLIQMPFFSAMYFAAQYTKGVSTSTFMGIDLGSRLVL 180

Query: 181 TVIIAILYFVQSWLSMQGVPDEQRQMKMTMYLMPIMMVFSISLPASVALYWFIGGIFS 240  
 T IIA LYF QSWLSM V +EQR+QMKTMMY MPIMM+FMS SLPA V LYW +GG FS  
 Sbjet: 181 TAIIAALYFFQSWLSMMAVSEEQREQMKMTMYTMPIMMIFMSFSLPAGVGLYWLVGFFS 240

15

20

25

30

```

Lipop: Possible site: 23      Crend: 6
McG: Discrim Score:          8.74
GvH: Signal Score (-7.5): -1.47
      Possible site: 16
>>> May be a lipoprotein
ALOM program    count: 4 value: -12.52 threshold: 0.0
  INTEGRAL      Likelihood ==-12.52    Transmembrane    60 - 76 ( 54 - 83)
  INTEGRAL      Likelihood = -3.66     Transmembrane   178 - 194 ( 177 - 196)
  INTEGRAL      Likelihood = -2.76     Transmembrane   140 - 156 ( 137 - 157)
  INTEGRAL      Likelihood = -2.60     Transmembrane   216 - 232 ( 213 - 232)
  PERIPHERAL    Likelihood = 0.74      235
modified ALOM score: 3.00

*** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.6010(Affirmative) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

```

37.9/63.7% over 193aa
Bacillus subtilis
EGAD|45886| hypothetical 30.7 kd lipoprotein in glng-ansr intergenic region precursor
Insert characterized
SP|P54544|YQJG_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN IN GLNQ-ANSR INTERGENIC REGION
PRECURSOR. Insert characterized
GP|1303958|dbj|BAA12613.1||D84432 YqjG Insert characterized
GP|2634823|emb|CAB14320.1||Z99116 similar to lipoprotein SpoIIIJ-like Insert
characterized
PIR|G69963|G69963 lipoprotein SpoIIIJ-like homolog yqjG - Insert characterized

ORF02470(478 - 1038 of 1530)
EGAD|45886|BS2384(63 - 256 of 275) hypothetical 30.7 kd lipoprotein in glng-ansr intergenic
region precursor {Bacillus subtilis}SP|P54544|YQJG_BACSU HYPOTHETICAL 30.7 KDA LIPOPROTEIN
IN GLNQ-ANSR INTERGENIC REGION PRECURSOR.GP|1303958|dbj|BAA12613.1||D84432 YqjG {Bacillus
subtilis}GP|2634823|emb|CAB14320.1||Z99116 similar to lipoprotein SpoIIIJ-like {Bacillus
subtilis}PIR|G69963|G69963 lipoprotein SpoIIIJ-like homolog yqjG - Bacillus subtilis
%Match = 13.0
%Identity = 37.9 %Similarity = 63.7
Matches = 72 Mismatches = 65 Conservative Sub.s = 49

```

55

60

65



-1725-

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]
Identities = 116/217 (53%), Positives = 168/217 (76%)

Query: 2  INWDAIFNLELAVKAFPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLRLRYLAN 61
I W+ IFN +LA+++FP VI+G+ YT+ +S V   G ++G F++L +MS  LLR+ A
Sbjct: 5  IQWEYIFNTKLAIESFPYVIKIGYGTLLISFVSMFAGTVIGLFISLARMSKLALLRWPAP 64

Query: 62  IHISLMRGIPLMVLLFLIYFGLPFFIGIQLDAVTASIVGFTMMSSAYISEIIRAALLAVDH 121
++IS MRG+P++V+LF++YFG P+IGI+  AVIA+++GF++ S+AYI+EI R+A+ +V+
Sbjct: 65  LYISFMRGVPILVILFIFYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEK 124

Query: 122  GQWEAARALGLKTPTIYRGIIIPQATRIALPSLSNVLLDMVKSSSLTAMITVPDIFNNAK 181
GQWEAA +LGL      RGII+PQ+  RIALP L+NVLLD++K+SSL AMITVP++  +AK
Sbjct: 125  GQWEAASSLGLSYWQTMRGIIIPQSIRIALPPLANVLLDLIKASSLAAMITVPELLOHAK 184

Query: 182  IVGGTYSDYMTAYILVALIYVICTLYAIIQDWWEKR 218
I+GG  DYMT YIL ALIYW IC++ A+ Q+  EK+
Sbjct: 185  IIGREFDYMTMYILTALIYWAICSAAVFQNILEKK 221
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4793> which encodes the amino acid sequence <SEQ ID 4794>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -6.79    Transmembrane 186 - 202 ( 184 - 205)
INTEGRAL    Likelihood = -5.84    Transmembrane 26 - 42 ( 21 - 43)
INTEGRAL    Likelihood = -4.78    Transmembrane 57 - 73 ( 56 - 84)
INTEGRAL    Likelihood = -1.59    Transmembrane 86 - 102 ( 86 - 103)

----- Final Results -----
bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12131 GB:Z99105 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]
Identities = 113/214 (52%), Positives = 157/214 (72%)

Query: 1  MINIPLMKDSLGLFVLSGLPYTLGISLSSFFTGLFLGLGLALLGRSRQPLIHYLVRAYISI 60
+ N L +S +V+ G+ YTL IS +S F G +GL ++L S+ L+ + + YIS
Sbjct: 10  IFNTKLAIESFPYVIKIGYGTLLISFVSMFAGTVIGLFISLARMSKLALLRWPAPLYISF 69

Query: 61  MRGVPMIVLVFLVLYFGLPYYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120
MRGVP++V+LF+LYFG PY G+E A+ A +GFS+ SAAYI+E+ RS+I +++KGQWEA
Sbjct: 70  MRGVPIILVILFIFYFGFPYIGIEFSAVTAALIGFSLNSAAYIAEINRSAISSVEKGQWEA 129

Query: 121  AKALGLPYALMVKKIILPQAFRIAPPLGNVIIDMVKSSSLAAMITVPDIFQNAKIIGGR 180
A +LGL Y  ++ IILPQ+  RIA+PPL NV++D++K+SSLAAMITVP++ Q+AKIIGGR
Sbjct: 130  ASSLGLSYWQTMRGIIIPQSIRIALPPLANVLLDLIKASSLAAMITVPELLOHAKIIGGR 189

Query: 181  EWDYMSMYILVAFIYWLI AFLRLERYQEFLNKLA 214
E+DYM+MYIL A IYW I +  +Q  LE K A
Sbjct: 190  EFDYMTMYILTALIYWAICSAAVFQNILEKKYA 223
```

An alignment of the GAS and GBS proteins is shown below.

```
Identities = 110/213 (51%), Positives = 156/213 (72%)

Query: 7  IFNLELAVKAFPSVIQGLPYTIGLSLVGFILGAIVGFFVALMKMSHFRLRLRYLANIHISL 66
+ N+ L +  V+ GLPYT+G+SL+ F G +G +AL+ S  L+ YL +IS+
Sbjct: 1  MINIPLMKDSLGLFVLSGLPYTLGISLSSFFTGLFLGLGLALLGRSRQPLIHYLVRAYISI 60
```

-1726-

Query: 67 MRGIPLMVLLFLIYFGLPFIGIQLDAVTASTVGFMTMSSAYISEIIRAALLAVDHGQWEA 126  
 MRG+P++V+LF++YFGLP+ G++L A+ + +GF+M+S+AYISE+ R+++ A+D GQWEA  
 Sbjct: 61 MRGVPMIVVLFVLYFGLPYGLELPALLCAYLGFSMVSAAYISEVFRSSIEAIDKGQWEA 120

Query: 127 ARALGLKTPTIYRGIIIPQATRIALPSLSNVLLDMVKSSSLTAMITVPDIFNNAKIVGGT 186  
 A+ALGL + + II+PQA RIA+P L NV++DMVKSSSL AMITVPDIF NAKI+GG  
 Sbjct: 121 AKALGLPYALMVKKIILPQAFRIAVPPLGNVIDMVKSSSLAAMITVPDIFQNAKIIGGR 180

Query: 187 YSDYMTAYILVALIYWVICTLYAIIQDWWEKRL 219  
 DYM+ YILVA IYW+I L Q++ E +L  
 Sbjct: 181 EWDYMSMYILVAFIYWLI AFLERYQEFL ENKL 213

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1553

A DNA sequence (GBSx1645) was identified in *S.agalactiae* <SEQ ID 4795> which encodes the amino acid sequence <SEQ ID 4796>. Analysis of this protein sequence reveals the following:

Possible site: 18  
 >>> May be a lipoprotein

----- Final Results -----  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12132 GB:Z99105 similar to amino acid ABC transporter  
 (binding protein) [Bacillus subtilis]  
 Identities = 127/276 (46%), Positives = 183/276 (66%), Gaps = 12/276 (4%)

Query: 3 KTILGLVGLSAMLTAACS--NGQSSKETTTWDNIKDGVLKVATPATLYPTSYYDDHK-- 58  
 K ++ + LAACS N SK+T W+ IK G + VAT TLYPTS+D  
 Sbjct: 8 KAVIFSFTMAFFLILAACSGKNEADSKDTGWEQIKDKGKIVVATSGTLYPTSYPHTDSDGS 67

Query: 59 -KLTGYEIDMMKAIKAKLKI KVFVEVGVAESFTSVD SGKVDVAVNNFDTTPERLKKYNF 117  
 KLTGYE++++ AK+L +KV+F E+G+ T+V+SG+VD A N+ D T +R +K+ F  
 Sbjct: 68 DKLTGYEVEVVREAAKRLGLKVEFKEMGIDGMLTAVNSGQVDAANDIDVT KDREKFAF 127

Query: 118 SQPYKYSVGGMIVRADGSSKITAKDLSDWKGKKAGGGAGTQYMKIAKQQAEPVIYDNVT 177  
 S PYKYS G IVR D S I K L D KGKKA G A T YM++A++ GA+ VIYDN T  
 Sbjct: 128 STPYKYSYGTAVRKDDL SGI--KTLKDLKGKKAAGAATTVMYEVARKYGAKEVIYDNAT 185

Query: 178 NDVYL RDVSTGR TDFIPNDYYTQVI AVKYVTQY PDIKVKM-GDVKYNPTEQGIVMSKKD 236  
 N+ YL+DV+ GR TD I NDYY Q +A+ +PD+ + + D+KY P +Q +VM K +  
 Sbjct: 186 NEQYLKDVANGRTDVILNDYYLQTLAL----AAFEDLNITIH PDIKYMPENKQALVMKSN 241

Query: 237 KSLKTKIDAAIKDMKKDGS LKKISEKYYAGODLTKE 272  
 +L+ K++ A+K+M KDGSL K+S++++ D++K+  
 Sbjct: 242 AALQKKMNEALKEMSKDGS LTKLSKQFFNKADVSKK 277

There is also homology to SEQ ID 1190.

SEQ ID 4796 (GBS183) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 2; MW 33kDa).

GBS183-His was purified as shown in Figure 199, lane 7.

-1727-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1554

A DNA sequence (GBSx1646) was identified in *S.agalactiae* <SEQ ID 4797> which encodes the amino acid sequence <SEQ ID 4798>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1514(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAF09821 GB:AE001885 6-aminohexanoate-cyclic-dimer hydrolase  
[Deinococcus radiodurans]  
Identities = 178/488 (36%), Positives = 265/488 (53%), Gaps = 17/488 (3%)

20 Query: 5 DATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSQYNEARQAQKYANESNA--- 61  
DA + Q ++ ++S++++ AI++ + NV++NAVV Y++ A+ + + A  
Sbjct: 54 DALDLAQLFRRGELSAEDMCTAAIHRAQVNVVALNAVVPYLYDQGLAQARATDAARARGE 113

25 Query: 62 ----PFAGVPILLKDLGQNGQLSTSGSQLFKHYHAKQTDYLVQSFEKLGFIILGRINT 117  
PFAGVP L+KD G G T G++ ++ + D LV+ ++ G + LG+TNT  
Sbjct: 114 QATGPFAGVPFLVKDFGSRLAGVPHTGGTRAYRDQIPEWDDLVRRWQAAGLLPLGKTNT 173

30 Query: 118 PEFQFKNISDQQLHGNVNLFPDHSRNAGSSSGGAAAVSSGMVPIAGSDGGGSIRIPAS 177  
PEF +++ +LHG P+D R GGSSGG+A+AV++G+VP+AGA DGGGSIRIPAS  
Sbjct: 174 PEFALMGVTEPELHGPTRNPDWDLGRTPGGSSGGSASAVAAGIVPLAGAGDGGGSIRIPAS 233

35 Query: 178 FNGLIGLKPSRGRIPVGPSSYRGWQGASSHFAITKSVRDTKRLLYLYLQSYQVES----PF 233  
GL GLKPSRGR+P G WQGA+ LT+SVRD+ LL Q + P  
Sbjct: 234 CCGFLGLKPSRGRVPCGDVGEPWQGAAVEHVLTRSVRDSAAALLDLEQGPDAGAALFLPS 293

40 Query: 234 PLKKLSKESLFEFSVSKPLKIAVLMDSPKTKVSSEAKAAIKEAADFLSQKGNHLELVEQ 293  
P + S+E E L+I PL V E AA++ AA L G+ +E V  
Sbjct: 294 PERPYSEEVGRE---PGRLRIGFSTAHLGRSVHPECVAAVQGAARLLES LGHEVEEVAL 350

45 Query: 294 PLDGIHSMKTYCMNSVETAAMFDDIEKSLGRSMEFSMDMELMTWAMYQSGQRVLA KDYSK 353  
P DG + + M+ ET A + +LGR SD+E +TW + Q G+ A D++  
Sbjct: 351 PWDGPALQAFLMLYFGETGASLAALRDTLGRPARASDVEAVTWLLGQLGRSYSAADFAA 410

50 Query: 354 LLDSDWQFAATMARFHENYDLILTAATNQAPFHHQFD---LDETQKQLRHMGEFSVSE 410  
SW+ A M RFH+NYDL+LT P G+ + L + + M +  
Sbjct: 411 ARASWNVHARAMGRFHQNYDLLLTPVLATPPLQIGELQPRGVQAALLRAAQQMDVSGLLR 470

55 Query: 411 QQDLIWKMFEDSMAWTPFTHQPNLTGQPSLAIPHLTKKGLPLGVQLTAAKGREDLLLA V 470  
+ + + D + P+T NLTGQP++++P H T +GLP+GVQ A RED+LL +  
Sbjct: 471 RSGQVDALATDILEKMPYTLANLTGQPAMSVPLHWTADGLPVGVQFVAPLAREDVLLRL 530

50 Query: 471 AELFEKEK 478  
A E+ +  
Sbjct: 531 AGQLEQAR 538

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4047> which encodes the amino acid sequence <SEQ ID 4048>. Analysis of this protein sequence reveals the following:

Possible site: 15  
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

60

-1728-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 277/484 (57%), Positives = 348/484 (71%), Gaps = 2/484 (0%)

Query: 1 MVFKDATAMVQAIKQHKISSQELVEQAIYKIEEQNVSVNAVVSQYNEARQAAKYANESN 60  
 M ++DATAM A++ + + ELV QAIYK ++ N ++NA+ S+++ A + AK + S  
 10 Sbjet: 1 MTYQDATAMAIAVQTGQTTPLELVQAIYKAKKLNPTLNATTSERFEAALEEAKQDFSG 60

Query: 61 APFAGVPILKDLGQNKQQLSTSGSQLFKHYHAKQTDYLVQSFEKLGFIILGRNTTPEF 120  
 PFAGVP+ LKDLGQ KG STSGS+LFK Y A +TD V+ E LGFIILGR+NTPEF  
 15 Sbjet: 61 LPFAGVPLFLKDLGQELKGHSSTSGSRLFKYQATKTDLFVKRLAELGFIILGRSNTTPEF 120

Query: 121 GFKNISDQQLHGNVNLPPDHSRNAGSSGGAAAVSSGMVPIAGASDGGGSIRIPASFNG 180  
 GFKNISD LHG VNLP D++RNAGSSGGAAA VSSG+ +A ASDGGGSIRIPASFNG  
 20 Sbjet: 121 GFKNISDSSLHGFPVNLPRDNRNAGSSGGAAALVSSGISALATASDGGGSIRIPASFNG 180

Query: 181 LIGLKPSRGRIPVGPSSYRGWQGASSHFALTKSVRDTKRLLYYLQSYQVESPFPLKKLSK 240  
 LIGLKPSRGR+PVGP SYR WQGAS HFALTKSVRDT+ LLYYLQ Q+ESPFPL L+K  
 25 Sbjet: 181 LIGLKPSRGRMPVGPSSYRSWQGASVHFALTKSVRDTNRLLYYLQMEQMESPFPLATLTK 240

Query: 241 ESLFEFSVSKPLKIAVLMDSPKTKVSSEAKAIAKEAADFLSQKGNHL-ELVEQPLDGIH 299  
 +S+++ S+ +PL IA + VS + A+++A +L ++G+ L EL E P++  
 30 Sbjet: 241 DSIYQ-SLQRPLTIAFYQRLSDGSPVSLDTAKALRQAVTWLREQHQLVELEEFFVNMT 299

Query: 300 SMKTYCMMNSVETAAMFDDIEKSLGRSMEFSDMELMTWAMYQSGQORVLAKDYSKLLDSWD 359  
 ++ Y +MNSVETAAMF DIE + GR M DME MTWA+YQSG+ + A YS++L WD  
 35 Sbjet: 300 VIRHYIIMNSVETAAMFADIEDTFGRPMTKDDMETMTWAIYQSGKDIPAWRYSQLQKWD 359

Query: 360 QFAATMARFHENYDLILTAAATNQAPFHHGQFDLDETLOKQLRHMGEFSVSEQQDLIWKMF 419  
 ++ATMA FHE YDL+LT TN PAP HG+ D L L FS EQ +L+ MF  
 40 Sbjet: 360 TYSATMASFHETYDLLLLTFTINTPAPKHGELVPDSKLMANLAQAEIFSSEEQFNLVETMF 419

Query: 420 EDSMAWTPFTHQPNLTGQPSLAIPHTLTKEGLPLGVQLTAAKGREDLLLAVAELFEKEKQ 479  
 S+A P+T PNLTGQP++++PT+ TKEGL +G+QL AAKGREDLLL +AE FE  
 45 Sbjet: 420 GKSLAINPYTALPNLTGQPAISLPYETKEGLSMGIQLIAAKGREDLLLGIAEQFEAAGL 479

Query: 480 FKGP 483  
 K P  
 50 Sbjet: 480 LKIP 483

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 45 vaccines or diagnostics.

### Example 1555

A DNA sequence (GBSx1647) was identified in *S.agalactiae* <SEQ ID 4799> which encodes the amino acid sequence <SEQ ID 4800>. This protein is predicted to be transcription elongation factor (greA). Analysis of this protein sequence reveals the following:

50 Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 55 bacterial cytoplasm --- Certainty=0.5003(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:CAB14674 GB:Z99117 transcription elongation factor [Bacillus subtilis]  
 Identities = 86/154 (55%), Positives = 114/154 (73%), Gaps = 1/154 (0%)

-1729-

Query: 3 EKTYPMQTQVEKDQLEKELEELKLVRRPEVVERIKIARSGDLSSENSEYDAAKDEQAFVEG 62  
 EK +PMT K +LE+ELE LK V+R EVVERIKIARS+GDLSSENSEYD+AK+EQAFVEG  
 Sbjct: 4 EKVPFMTAEGKQKLEQLELYLKTVKRKEVVERIKIARSPGDLSSENSEYDSAKEEQAFVEG 63

Query: 63 QIQILETKIRYAEIIDS DAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGKIS 122  
 ++ LE IR A+II+ D + V +GKTV E+ D+++Y IVG+A AD F GKIS  
 Sbjct: 64 RVTTLNMIIRNAKIIEDDG-GSNVVGLGKTVTFVELPDGDEESYTTIVGSAEADPFEGKIS 122

Query: 123 NESPIAHALIGKKTGDLATIESPAGSYQVEIISV 156  
 N+SPIA +L+GKK + T+++P G V+I+ +  
 Sbjct: 123 NDSPIAKSLGKKVDEEVTQTTPGGEMLVKIVKI 156

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4801> which encodes the amino acid  
 sequence <SEQ ID 4802>. Analysis of this protein sequence reveals the following:

Possible site: 27  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4434 (Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/160 (90%), Positives = 149/160 (92%)

Query: 1 MAEKTYPMTQVEKDQLEKELEELKLVRRPEVVERIKIARSGDLSSENSEYDAAKDEQAFV 60  
 MAEKTYPMT EK+QLEKELEELKLVRRPE+VERIKIARSGDLSSENSEYDAAKDEQAFV  
 Sbjct: 17 MAEKTYPMTLTEKEQLEKELEELKLVRRPEIVERIKIARSGDLSSENSEYDAAKDEQAFV 76

Query: 61 EGQIQILETKIRYAEIIDS DAVAKDEVAIGKTVLVQEVGTNDKDTYHIVGAAGADIFSGK 120  
 EGQI LETKIRYAEIIDS DAVAKDEVAIGKTV+VQEVGT DKDTYHIVGAAGADIFSGK  
 Sbjct: 77 EGQISTLETKIRYAEIIDS DAVAKDEVAIGKTVIVQEVGTTDKDTYHIVGAAGADIFSGK 136

Query: 121 ISNESPIAHALIGKKTGDLATIESPAGSYQVEIISVEKTN 160  
 ISNESPIA ALIGKKTGD IESPA +Y VEIISVEKTN  
 Sbjct: 137 ISNESPIAQLIGKKTGDKVRIESPAATYDVEIISVEKTN 176

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### Example 1556

A DNA sequence (GBSx1648) was identified in *S.agalactiae* <SEQ ID 4803> which encodes the amino  
 acid sequence <SEQ ID 4804>. This protein is predicted to be aminodeoxychorismate lyase-like protein.  
 Analysis of this protein sequence reveals the following:

Possible site: 58  
 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -13.64 Transmembrane 238 - 254 ( 230 - 260)

----- Final Results -----

bacterial membrane --- Certainty=0.6456 (Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein  
 [Streptococcus thermophilus]  
 Identities = 135/210 (64%), Positives = 171/210 (81%)

-1730-

Query: 373 KTTSTPYKADDFLKLVDQETFIKMMVAKYPNLLGSLPDKSKAIYQLEGYLFPATYNYKD 432  
 K +ST K DFLKL++D+ FI KM AKYP LL +LP+ + A Y LEGYLFPATYN + D  
 Sbjct: 5 KHSSTGLKEKDFLKLMDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPATYNIHDD 64

5 Query: 433 TTLEGLVEDMISTMNTKMAPYNTIKAKNMSVNDVLTSSLVEKEGSTDEDRRKIASVFY 492  
 TT+E L E+M+ TM+T ++PYY TI + N +VN++LTL+SLVEKEG+TD+DR+ IASVFY  
 Sbjct: 65 TTVESLAEEMLFMDTHLSPPYATILSSNHNVEILTASLVEKEGATDDDRKNIASVFY 124

10 Query: 493 NRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNIYNTGLMPGPVDSPSI 552  
 NRL++ ALQSNIA+LY +GKLG +T+L EDA I+T+I SPYN Y + GLMPGPVDSPS+  
 Sbjct: 125 NRINSDMALQSNIAVLVYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDSPSL 184

Query: 553 SAIEATIKPASTDYLYFVADVKTGNVYAK 582  
 SAIEA I P+ST Y+YFADV TGNVY+A+  
 15 Sbjct: 185 SAIEAVINPSSTKMYFVADVSTGNVYFAE 214

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4805> which encodes the amino acid sequence <SEQ ID 4806>. Analysis of this protein sequence reveals the following:

Possible site: 59

20 >>> Seems to have no N-terminal signal sequence  
 INTEGRAL Likelihood = -7.91 Transmembrane 161 - 177 ( 155 - 183)

25 ----- Final Results -----  
 bacterial membrane --- Certainty=0.4163(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

30 >GP:AAF77615 GB:AF151720 aminodeoxychorismate lyase-like protein  
 [Streptococcus thermophilus]  
 Identities = 135/212 (63%), Positives = 161/212 (75%)

35 Query: 295 KTKKAKTPFNEKDFLDLVTDEAFIQDMVKRYPKLLATIPTKEKAIYRLEGYLFPATYNY 354  
 K K + T EKDFL L+ D+AFI M +YP LLA +P A Y LEGYLFPATYN +  
 Sbjct: 3 KGHSSSTGLKEKDFLKLMDDAFITKMKAKYPTLLANLPNSTDAKYVLEGYLFPATYNIH 62

Query: 355 KETTMRELVEDMLAAMDATLVPPYDKIAASGKTVNEVLTLASLVEKEGSTDDRRQIASV 414  
 +TT+ L E+ML MD L PYY I +S VNE+LTLASLVEKEG+TDDR+ IASV  
 40 Sbjct: 63 DDTTVESLAEEMLFMDTHLSPPYATILSSNHNVEILTASLVEKEGATDDDRKNIASV 122

Query: 415 FYNRLNSGMALQSNIAILYAMGKLGKETTLEADATIDTTINSFYNIYNTGLMPGPVASS 474  
 FYNRLNS MALQSNIA+LY +GKLG++TTL EDATIDT I+SPYN Y + GLMPGPV S  
 45 Sbjct: 123 FYNRLNSGMALQSNIAVLVYVLGKLGQETTLKEDATIDTNIDSPYNDYVHKGLMPGPVDS 182

Query: 475 GVSIEATLNPASTDYLYFVANVHTGEVYAK 506  
 +SAIEA +NP+ST Y+YFVA+V TG VY+A+  
 Sbjct: 183 SLSAIEAVINPSSTKMYFVADVSTGNVYFAE 214

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 310/603 (51%), Positives = 403/603 (66%), Gaps = 86/603 (14%)

Query: 1 MTEFNDDQHSNHDQKSFKEQILAELEEANRLRKLREEELYQKEQEAARRTAQLMADY 60  
 +T+F D + Q+SFKEQILAELE+AN++RK +EEL+  
 55 Sbjct: 3 LTDFKDKDQQDQ-QRSFKEQILAELEKANQIRKEKEELF----- 41

Query: 61 EAQRLKDEREARAKALETKQRLLEEQEKARIEAKLLAEAAAREEERRQAEQALASQEEQVIN 120  
 ++ LE +E AR A+L AE R++ A Q+E + +  
 60 Sbjct: 42 -----QKELEAKEAARRTAQLYAEYKRQD-----AFQKESIAH 74

Query: 121 QGMEPSRELDSGSKSSEFRTTENVPDIDLKADKTDVATAVPNQETERIFLVRATDIPTG 180  
 +T ++ +A K V T+ + T + +E  
 Sbjct: 75 NN-----KTAKH-----FQAIKGAVMTSEALKPT-----LLSEK 103

-1731-

Query: 181 ENVKLGEISELEPVAKEPIRVEDLSKEEEGIALSAKNKHNRER--RQKADNVAKRIAR 237  
 EN L ++ A E +++ + +E + L+ + H+ R + RQ+ + AK+I+  
 Sbjct: 104 ENSSLKTTNKRVRVQANE---LQETASKESQVPLTIEKGHSVRRKLSKRQQTERAAKKIST 160

5 Query: 238 ILISIIILVLLLTAFVGYRFVDSAIKPVDSNSNKFVQVEIPIGSGNKLIGQILEKAGVIK 297  
 +LIS II+ LL G +V SA+ PVD NS+ FVQVEIP GSGNKLIGQIL+K G+IK  
 Sbjct: 161 VLISSIIITLLAVTLAGAGYVYSALNPVDKNSDAFVQVEIPSGSGNKLIGQILQKKGLIK 220

10 Query: 298 SATVFNYYSKFKNYSNFQSGYYNLKKSMTLDQIAAELEKGGTAEPTKPALGKILITEGYT 357  
 ++TVF++Y+KFKN++NFQSGYYNL+KSM+L++IA+ L++GGTAEPTKP+LGKILI EGYT  
 Sbjct: 221 NSTVFSFYTKFKNFTNFQSGYYNLQKSMSLEEIASALQEGGTAEPTKPSLGKILIPEGYT 280

15 Query: 358 IKQIAKAIESN-KIDTKTTSTPYKADDFLKLVDQETFIKKMVAKYPNLLGSLPDKSKAIY 416  
 IKQIAKA+E N K TK TP+ DFL LV DE FI+ MV +YP LL ++P K KAIY  
 Sbjct: 281 IKQIAKAVEHNSKGKTKKAKTPFNEKDFLDLVTDFAFIQDMVKRYPKLLATIPTKEKAIY 340

20 Query: 417 QLEGYLFPPATYNYKYDTTLEGLVEDMISTMTKMAPYYNTIKAKNMSVNDVLTLSLVEK 476  
 +LEGYLFPPATYNYK+TT+ LVEDM++ M+ + PYY+ I A +VN+VLT+SLVEK  
 Sbjct: 341 RLEGYLFPPATYNYKETTMRELVEDMLAAMDATLVPPYDKIAASGKTVNEVLTASLVEK 400

25 Query: 477 EGSTDEDRRKIASVFYNRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNI 536  
 EGSTD+DRR+IASVFYNRL++G ALQSNIAILYAMGKLG+KT+LAEDA I+T+I SPYNI  
 Sbjct: 401 EGSTDDRRQIASVFYNRLNSGMALQSNIAILYAMGKLGDKTSLAEDATIDTTINSFYNI 460

30 Query: 537 YNTNTGLMPGPVDSPSISAEATIKPASTDYLYFVADVKTGNVYYAKDFETHKANVEKYIN 596  
 YNTNTGLMPGPV S +SAIEAT+ PASTDYLYFVA+V TG VYYAK FE H ANVEKY+N  
 Sbjct: 461 YNTNTGLMPGPVASSGVSAIEATLNPASTDYLYFVANVHTGEVYYAKTFEEHSANVEKYVN 520

Query: 597 SQI 599  
 SQI  
 Sbjct: 521 SQI 523

A related GBS gene <SEQ ID 8843> and protein <SEQ ID 8844> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 8  
 McG: Discrim Score: -17.88  
 GvH: Signal Score (-7.5): -3.51  
 Possible site: 58  
 >>> Seems to have no N-terminal signal sequence

40 ALOM program count: 1 value: -13.64 threshold: 0.0  
 INTEGRAL Likelihood = -13.64 Transmembrane 238 - 254 ( 230 - 260)  
 PERIPHERAL Likelihood = 5.78 285  
 modified ALOM score: 3.23

45 \*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.6456(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00931(1417 - 2046 of 2400)  
 GP|8574530|gb|AAF77615.1|AF151720\_1|AF151720(5 - 214 of 214) aminodeoxychorismate lyase-  
 55 like protein {Streptococcus thermophilus}  
 %Match = 17.5  
 %Identity = 64.3 %Similarity = 81.4  
 Matches = 135 Mismatches = 39 Conservative Sub.s = 36

60 1236 1266 1296 1326 1356 1386 1416 1446  
 NYYSKFKNYSNFQSGYYNLKKSMTLDQIAAELEKGGTAEPTKPALGKILITEGYTIKQIAKAIESNKIDTKTTSTPYKAD  
 . . . . .  
 AKKGKHSSTGLKEK  
 10

65

-1732-

1476 1506 1536 1566 1596 1626 1656 1686  
 DFLKLVQDEFIKMVAKYPNLLGSLPDKSKAIYQLEGYLFPPATYNYKDTTLEGLVEDMISTMTKMAPYYNTIKAKNM  
 |||||::|: || || |||| || :||: | | ||||| ||||| : |||: | |::| ||: || : |  
 5 DFLKLMKDDAFITKMKAKYPTLLANLPNSTDAKVLEGYLFPPATYNIHDDTTVESLAEEMLFMTDTHLSPYYATILSSNH  
 30 40 50 60 70 80 90

1716 1746 1776 1806 1836 1866 1896 1926  
 SVNDVLTLSLVEKEGSTDEDRRKIASVFYNRLSAGQALQSNIAILYAMGKLGDKTSLAEDAQINTSIKSPYNIYNTTGL  
 :||::|||:|||||||:|:|: ||||| ||||| : |||| :|: || |::| |||| | : ||  
 10 NVNEILTASLVEKEGATDDDRKNIASVFYNRLNSDMALQSNIAVLVYLGKLGQETTLKEDATIDTNIDSPYNDYVHKGL  
 110 120 130 140 150 160 170

1956 1986 2016 2046 2076 2106 2136 2166  
 MPGPVDSPLSAIEATIKPASTDYLYFVADVKTGNVYAKDFETHKANVEKYINSQIN\*AYKHGASHHVYIFDLKK\*KEK  
 ||||| ||||| :||| | |::| ||||| ||||| :|:  
 15 MPGPVDSPLSAIEAVINPSSTKYMYFVADVSTGNVYFAE  
 190 200 210

SEQ ID 8844 (GBS370) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 6; MW 70kDa).

20 GBS370-His was purified as shown in Figure 209, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1557

25 A DNA sequence (GBSx1649) was identified in *S.agalactiae* <SEQ ID 4807> which encodes the amino acid sequence <SEQ ID 4808>. Analysis of this protein sequence reveals the following:

Possible site: 53  
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.0183(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 10077> which encodes amino acid sequence <SEQ ID 10078> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA98889 GB:Z74367 ORF YDR071c [Saccharomyces cerevisiae]  
 Identities = 52/174 (29%), Positives = 81/174 (45%), Gaps = 18/174 (10%).

40 Query: 27 MSMIIRNGCLEDLQQVISIEQINFSEAEAAASKKAMQERLTIMTDT---FLVAEINGR--- 80  
 + M IR +EDL+Q++++E F E AS++ + RL + + EI G+  
 Sbjct: 10 LHMYIRPLIIEDLKQILNLESQGFPPNERASEEIIISFRLINCPCLCSGLFIREIEGKEVK 69

45 Query: 81 ---LAGYIEGPFVIKGRYLTDDLPHKVSEFPVRVGGFIGITSLSIHPDFKGGIGTALLAA 137  
 L G+I G I Y+T + K+ V IGI S+ I P+++ + + T LL  
 Sbjct: 70 KETLIGHIMGTKIPHEYITIESMGKLQ---VESSNHIGIHSVVIKPEYQKKNLATLLLT 126

Query: 138 MKDLVVQSQE-RDGISLTCHDDLISFYEMNGFKDEGES-----DSKHGGSLWYNM 185  
 + +QE + I L H+ LI FYE GFK E+ D W +M  
 50 Sbjct: 127 YIQKLSNQEIGNKIVLIAHEPLIPFYERVGFKIIAENTNVAKDKNFAEQKWIDM 180

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4809> which encodes the amino acid sequence <SEQ ID 4810>. Analysis of this protein sequence reveals the following:

55 Possible site: 15  
 >>> Seems to have no N-terminal signal sequence

-1733-

## ----- Final Results -----

5           bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>  
           bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 87/159 (54%), Positives = 117/159 (72%), Gaps = 1/159 (0%)

10   Query: 29 MIIRNGCLEDLQQVISIEQINFSEAEAAASKKAMQERLTIMTDTFLVAEINGRLAGYIEGP 88  
           M+IR       DL+ + +IE NFS EA ++ ++E + ++ DTFLVA I+ + GYIEGP  
   Sbjct: 1 MLIRQVQGSDELVIATIESDNFSPQEAATRAVLEEHIRLIPDTFLVALIDQEIYGYIEGP 60

15   Query: 89 VIKGRYLTDDLPHKVSEFPVRVGGFIGITSLSIHPDFKQGIGTALLAAMKDLVVSQERD 148  
           V+       L D LFH V++ P + GG+I ITSLSI F+ QG+GTALLAA+KDLVV+Q+R  
   Sbjct: 61 VVTTPILEDLSLFHGVTKNP-KTGGYIAITSLSIKHFQQQGVGTALLAALKDLVVAQQRT 119

20   Query: 149 GISLTCHDDDISFYEMNGFKDEGESDSKHGGSLWYNMIW 187  
           G+ LTCHD LIS+YEMNGF ++G S+S+HGG+LWY MIW  
   Sbjct: 120 GLILTCHDYLLISYEMNGFINQGISSESQHGGTLWYQMIW 158

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

**Example 1558**

25   A DNA sequence (GBSx1650) was identified in *S.agalactiae* <SEQ ID 4811> which encodes the amino acid sequence <SEQ ID 4812>. This protein is predicted to be udp-n-acetylmuramate--alanine ligase (murC/ddlA). Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have a cleavable N-term signal seq.

30   INTEGRAL       Likelihood = -2.60   Transmembrane   272 - 288 ( 270 - 288)

## ----- Final Results -----

35           bacterial membrane --- Certainty=0.2041(Affirmative) < succ>  
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine  
           ligase [Bacillus subtilis]

40   Identities = 238/432 (55%), Positives = 315/432 (72%), Gaps = 3/432 (0%)

Query: 5 YHFIGIKGSGMSALALMLHQMGNHVQGSDDVKYYFTQRGLEQAGVTILPFPSPNNISEDLE 64  
           YHF+GIKG+GMS LA +LH G+ VQGSDD++K+ FTQ LE+ +TILPFS NI +  
   Sbjct: 4 YHFVGIKGTGMSPLAQILHDNGYTVQGSDDIEKFIFTQTALEKRNITILPFSANIKPGMT 63

45   Query: 65 IIAGNAFRPDNNEELAYVIEKGYQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGLLAHV 124  
           +IAGNAF PD + E+ + +G RYH+FLGD+M++FTS+ V GAHGKTSTTGLLAHV  
   Sbjct: 64 VIAGNAF-PDTHPEIEKAMSEGIPVIRYHKFLGDYMKKFTSVAVVGAGHKTSTTGLLAHV 122

50   Query: 125 LKNITDTSFLIGDGTGRGSANANYFVFEADEYERHFMPYHPEYSIITNIDFDHPDYFTGL 184  
           ++N TSFLIGDGTG+G+ N+ YFVFEA EY RHF+ Y P+Y+I+TNIDFDHPDYF+ +  
   Sbjct: 123 IQNAKPTSFLLIGDGTGQGNENSEYFVFEACEYRRHFLSYQPDYAIMNIDFDHPDYFSSI 182

55   Query: 185 EDVFNAPNDYAKQVQKGLFTYGEDPKLHEITSEAPIYYYGFEDSNDFIKDIITRTVNGSD 244  
           +DVF+AF + A QV KG+ G+D L+I + P+ YYG + NDF A++I ++ G+  
   Sbjct: 183 DDVFDFAQEMALQVNKGIIACGDDEHLPKIHANVPVYYYGTGEENDFQARNIVKSTEGTT 242

60   Query: 245 FKVFFYNQEEIGQFHVHPAYGKHNLNATAVIANLYIMGIDMALVAEHLKTFSGVKRRFTEK 304  
           F VF F++PAYG HN+LN+ AVIA + ID +++ LK+F GVKRRF EK  
   Sbjct: 243 FDFVFRNTFYDTFYIPAYGHHNVLNLSLAVIALCHYEEIDSSIIKHALKSFGGVKRRFNEK 302

-1734-

Query: 305 IIDDTVIIDDFAHHPTEIATLDAARQKYPskeIVAIFQPHFTFTRTIALLDEFahalSQA 364  
 + D V+IDD+AHHPTEI T++AARQKYP +EIVA+QPHFTFTRT LDEFA +LS A  
 Sbjct: 303 QLGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHFTFTRTQQFLDEFAESLSGA 362

Query: 365 DSVYLAQIYGSAREVDNGEVVKVEDLAAKIVKHSDDLVTVENVSPLLNDHNAVYVFMGAGDI 424  
 D VYL I+GSARE + G++ + DL KI ++ L+ ++ S L HD AV +FMGAGDI  
 Sbjct: 363 DCVYLCDFGSARE-NAGKLTIGDLQGI-HNAKLIEEDDTSVLKAHDKAVLIFMGAGDI 420

Query: 425 QLYERSFEELLA 436  
 Q Y R++E ++A  
 Sbjct: 421 QKYMRAYENVMA 432

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4813> which encodes the amino acid  
 sequence <SEQ ID 4814>. Analysis of this protein sequence reveals the following:

Possible site: 31  
 >>> Seems to have a cleavable N-term signal seq.  
 INTEGRAL Likelihood = -4.57 Transmembrane 271 - 287 ( 269 - 288)

----- Final Results -----  
 bacterial membrane --- Certainty=0.2826(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC00294 GB:AF008220 putative UDP-N-acetylmuramate-alanine  
 ligase [Bacillus subtilis]  
 Identities = 236/431 (54%), Positives = 310/431 (71%), Gaps = 2/431 (0%)

Query: 5 YHFIGIKGSGMSALALMLHQMGMHKGSDVEKYFTQGLEQAGITILPFSNITPDME 64  
 YHF+GIKG+GMS LA +LH G+ VQGS+EK+ FTQ LE+ ITILPFS +NI P M  
 Sbjct: 4 YHFVGIKGTGMSPLAQILHDNGYTVQGSIDIEKFIFTQTALEKRNITILPFSANIKPGMT 63

Query: 65 LIVGNAFRENNKEVAYALRHQIPFKRYHDFLGDFMKSFSISFAVAGHGKTSTTGLLSHVL 124  
 +I GNAF + + E+ A+ IP RYH FLGD+MK F S AV GAHGKTSTTGLL+HV+  
 Sbjct: 64 VIAGNAFPDTHPEIEKAMSEGIPVIRYHKFLGDYMKKFTSVAVTGANGKTSTTGLLAHVI 123

Query: 125 KNITDTSYILIGDGTGRGSANAQYFVFESDEYERHFMFYHPEYSIITNIDFDHPDYFTGIA 184  
 +N TS+LIGDGTG+G+ N++YFVFE+ EY RHF+ Y P+Y+I+TNIDFDHPDYF+ I  
 Sbjct: 124 QNAKPTSFLIGDGTQGNENSEYFVFEACEYRRHFLSYQPDYAIMTNIDFDHPDYFSSID 183

Query: 185 DVRAFNDYAKQVKKALFVYGEDDELKKIEAPAPIYYYGFEEGNDFIAYDITRTNGSDF 244  
 DV +AF + A QV K + G+D+ L KI A P+ YYG E NDF A +I ++T G+ F  
 Sbjct: 184 DVFDFAQEMALQVNGIACGDDEHLPKIHANVPVYYGTGEENDFQARNIVKSTEGTTF 243

Query: 245 KVKHQGEVIGQFHVPAVGKHNILNATAVIANLNFVAGIDMALVADHLKTFSGVKRRFTEKI 304  
 V + F++PAYG HN+LN+ AVIA ID +++ LK+F GVKRRF EK  
 Sbjct: 244 DVFVRNTFYDTFYIPAYGHHNVLNSLAVIALCHYEIDSSIHKHALKSFGGVKRRFNEKQ 303

Query: 305 INDITIIIDFAHHPTEIVATIDAARQKYPskeIVAIFQPHFTFTRTIALLEDFAALNEAD 364  
 + D ++IDD+AHHPTEI TI+AARQKYP +EIVA+QPHFTFTRT L++FA +L+ AD  
 Sbjct: 304 LGDQVLIDDYAHHPTEIKVTIEAARQKYPDREIVAVFQPHFTFTRTQQFLDEFAESLSGAD 363

Query: 365 SVYLAQIYGSAREVDKGEVVKVEDLAAKIIKPSQVVTVENVSPLLNDHNAVYVFMGAGDIQ 424  
 VYL I+GSARE + G+++ + DL K I ++++ ++ S L HD AV +FMGAGDIQ  
 Sbjct: 364 CVYLCDFGSARE-NAGKLTIGDLQGI-HNAKLIEEDDTSVLKAHDKAVLIFMGAGDIQ 421

Query: 425 LYEHSFEELLA 435  
 Y ++E ++A  
 Sbjct: 422 KYMAYENVMA 432

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/443 (83%), Positives = 406/443 (91%), Gaps = 1/443 (0%)

-1735-

Query: 1 MSKTYHFIGIKGSGMSALALMLHQMGNVQGSVDVKYYFTQRGLEQAGVTILPFSPNNIS 60  
 MSKTYHFIGIKGSGMSALALMLHQMGNH VQGSVDV+KYYFTQRGLEQAG+TILPFS +NI+  
 Sbjct: 1 MSKTYHFIGIKGSGMSALALMLHQMGNHVKVQGSVDVEKYYFTQRGLEQAGITILPFSEDNIT 60

5 Query: 61 EDLEIIAGNAFRPDNNEELAYVIEKGYQFKRYHEFLGDFMRQFTSLGVAGAHGKTSTTGL 120  
 D+E+I GNAFR +NN+E+AY + FKRYPH+FLGDFM+ F S VAGAHGKTSTTGL  
 Sbjct: 61 PDMELIVGNAFR-ENNKEVAYALRHQIPFKRYHDFLGDFMKSFISFAVAGAHGKTSTTGL 119

10 Query: 121 LAHVLKNITDTSFLIGDGTGRGSANANYFVFEADYERHFMYPHPEYSIITNIDFDHPDY 180  
 L+HVLKNITDTS+LIGDGTGRGSANA YFVFE+DEYERHFMYPHPEYSIITNIDFDHPDY  
 Sbjct: 120 LSHVLKNITDTSYLIGDGTGRGSANAQYFVFESEYERHFMYPHPEYSIITNIDFDHPDY 179

15 Query: 181 FTGLEDVFNAFNDYAKQVQKGLFIYGEDPKLHEITSEAPIYYYGFEDSNDFIKADITRTV 240  
 FTG+ DV NAFNDYAKQV+K LF+YGED +L +I + APIYYYGFE+ NDFIA DITRT  
 Sbjct: 180 FTGIADVRNAFNDYAKQVKALFVYGEDDELKKIEAPAPIYYYGFEEGNDFIAYDITRTT 239

20 Query: 241 NGSDFKVFYNQEEIGQFHVPAVGKHNILNATAVIANLYIMGIDMALVAEHLKTFSGVKRR 300  
 NGSDFKV + E IGQFHVPAVGKHNILNATAVIANL++ GIDMALVA+HLKTFSGVKRR  
 Sbjct: 240 NGSDFKVKHQGEVIGQFHVPAVGKHNILNATAVIANLFAVAGIDMALVADHLKTFSGVKRR 299

25 Query: 301 FTEKIIDDVTIIDDFAHHPTEIATLDAARQKYSKEIVAIFQPHTFTRTIALLDEFABA 360  
 FTEKII+DT+IIDDFAHHPTEI+AT+DAARQKYSKEIVAIFQPHTFTRTIALL++FA A  
 Sbjct: 300 FTEKIINDTIIIDDFAHHPTEIVATIDAARQKYSKEIVAIFQPHTFTRTIALLEDFAA 359

30 Query: 361 LSQADSVYLAQIYGSAREVDNGEVKVEDLAAKIVKHSDLVTVENVSPLLNDHNAVYVFMG 420  
 L++ADSVYLAQIYGSAREVD GEVKVEDLAAKI+K S +VTVENVSPLL+HDNAVYVFMG  
 Sbjct: 360 LNEADSVYLAQIYGSAREVDKGEVKVEDLAAKIIKPSQVVTVENVSPLLHDHNAVYVFMG 419

Query: 421 AGDIQLYERSFEELLANLTKNTQ 443  
 AGDIQLYE SFEELLANLTKN Q  
 Sbjct: 420 AGDIQLYEHSEELLANLTKNNQ 442

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 24 (lane 11; MW 49kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 8; MW 74kDa), Figure 33 (lane 8; MW 74kDa) and Figure 37 (lane 3; MW 74kDa).

The GBS157-GST fusion product was purified (Figure 112A; see also Figure 200, lane 3) and used to immunise mice (lane 1+2 product; 19.5µg/mouse). The resulting antiserum was used for Western blot (Figure 112B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 4812 (GBS157) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 11-13; MW 74kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1559

A DNA sequence (GBSx1651) was identified in *S.galactiae* <SEQ ID 4815> which encodes the amino acid sequence <SEQ ID 4816>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1980(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1736-

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4817> which encodes the amino acid sequence <SEQ ID 4818>. Analysis of this protein sequence reveals the following:

5      Possible site: 25  
      >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----  
          bacterial cytoplasm --- Certainty=0.2731(Affirmative) < succ>  
 10        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 80/201 (39%), Positives = 126/201 (61%), Gaps = 9/201 (4%)

15      Query: 7    RFPLIADDEPVMSPVLKMNLYDNEDLINNIRDFYQEKTYQSMVKSNYEHEEISHPKVIEN 66  
                 +FPL+AD    + P +M LY+NEDLI NIR +YQ+K Y + ++    EE +  
      Sbjet: 5    QFPLVADGIAISDPKQMALYENEDLITNIRGYQDKEYDDIARN-----EEFTAKATSRQ 60

20      Query: 67    DPVPPQ--SFVKKATELSKSRQEAQRSVREKQAYYAKQEFKAPSKEAFQQQLKATVPKK 124  
                 P    + S +K    + ++RQ+AK+ ++EKQAY AK+    P + + +QQ    + P +  
      Sbjet: 61    TPSSKRFCSDNDEKHHYVKEARQKAKQDLKEKQAYLAKEMAYVPKQVSKKQPADSSPSQ 120

25      Query: 125    QTQRKVTLSHLSDRLQQESYILAEIPIIFQEPDNTPNP-KTKKNNFDLKRQVYNKQD 183  
                 +    + TE+S    + +L Q++YILAE+P ++EP N P    TTKNN+DFLK SQ+YN ++  
      Sbjet: 121    K--QATTEMSRFTKKLHQDNYILAEPLKEYKEPKNLPPQGTTKKNNYDFLKSSQIYNNKE 178

30      Query: 184    NQFHKERAKAQELNLTFRKDI 204  
                 +    +E+    AQELNL+RF+D+  
      Sbjet: 179    MRQQREKTIAQELNLSRFEDL 199

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1560

35      A DNA sequence (GBSx1652) was identified in *S.agalactiae* <SEQ ID 4819> which encodes the amino acid sequence <SEQ ID 4820>. Analysis of this protein sequence reveals the following:

     Possible site: 29  
      >>> Seems to have no N-terminal signal sequence

40      ----- Final Results -----  
          bacterial cytoplasm --- Certainty=0.4959(Affirmative) < succ>  
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

45      No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

#### Example 1561

50      A DNA sequence (GBSx1653) was identified in *S.agalactiae* <SEQ ID 4821> which encodes the amino acid sequence <SEQ ID 4822>. This protein is predicted to be SNF. Analysis of this protein sequence reveals the following:

-1737-

Possible site: 28

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.32 Transmembrane 743 - 759 ( 743 - 759)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1128(Affirmative) &lt; succ&gt;

bacterial outside --- Certainty=0.0000(Not Clear) &lt; succ&gt;

bacterial cytoplasm --- Certainty=0.0000(Not Clear) &lt; succ&gt;

10 The protein has homology with the following sequences in the GENPEPT database.

&gt;GP:CAA67095 GB:X98455 SNF [Bacillus cereus]

Identities = 259/678 (38%), Positives = 406/678 (59%), Gaps = 21/678 (3%)

15 Query: 369 QNEILLQMVFDYGNDLTVHNRQEQLEQLTFASHFKHEEKVFKLLEKYGFAPHFSTSHPAYS 428  
 +N +L + F YGN + ++ + F K E+++ ++ + FA + ++

Sbjct: 388 KNRLLAGLEFHYGNVVINPLEEDGQPSVFNREKKEKEILDIMSESAFAKT-EGGYFMHN 446

20 Query: 429 AQELYDFYTYMLPQFKMGTV--SLSAKLESYRLIERPQIDIEAKGSL--LDISDFDSDL 484  
 + Y+F +++P K + + + + KL ++ P I + K + L FD +

Sbjct: 447 EEAENFLYHIVPTLKGVLDIYATTAIKLRIHKGDTAPLIRVRKERIDWLSFRFDIKGI 506

25 Query: 485 LENDVDQALVALFDNPNPYFVNKSGQLVIFD-EETKKVSATLQ--GLRARRAKNGHIELDN 541  
 E ++ L AL + Y+ +G L+ + +E +++ ++ G+R + +

Sbjct: 507 PEAEIKGVLAALAEKRYRRLANGSLLSLESKEFNEINQFVKESGIRKEFLHGEENVNPL 566

Query: 542 IAAFQSELSEFANQDNVSFSQHFIQIEDLRHPEKFK--IPGLSVSASLRDYQLTGVRWLS 599  
 I + + + +S + L+E +++P+K K +P ++ A +R+YQ+ G W+

Sbjct: 567 IRSVKWMNGLHEGNVLSLDESVDLVEIQNPKKLKFTVPP-TLHAVMREYQVYGFEMMK 625

30 Query: 600 MLDHYGFAGILADDMLGKTLQITISFLSTKLT--RDSR--VLILSPSSLIYNWQDEFHKEF 655  
 L +Y F GILADDMLGKTLQ+I+++ + L R+ + +L++SPSSL+YNW E KF

Sbjct: 626 TLAYYRFGGILADDMLGKTLQSIAYIDSVLPEIREKKLPILVSPSSLVINWVSELKKF 685

35 Query: 656 APDVDVAVAYGSKIRDEIIAE--RHQVIITSYSSFRQDFETYSEGNYDYILDEAQVMK 713  
 AP + +A G++ R +I+ + V+ITSY R+D +Y+ + L LDEAQ K

Sbjct: 686 APHIRAVIADGNQTERRKILKDVAEFDVVITSYPLLRDVRSYARP-FHTLFLDEAQAFK 744

40 Query: 714 NAQTKIAHSLRSFEVKNCFALSGTPIENKLEIWSIFQIILPGLLPKGKEFLKLNPKQVA 773  
 N T+ A +++ + + F L+GTP+EN L E+WSIF ++ P LLPG+KEF L + +A

Sbjct: 745 NPTTQTARAVKTIQAEYRFGLTGTPVENSLEELWSIFHVVFPELLPGRKEFGDLRREDIA 804

45 Query: 774 RYIKPFVMRRRKEEVLPELPDLIEMNYPNEMTDSQKVIYLAQLRQI-QESIQHSSDADLN 832  
 +KPFV+RR KE+VL ELPD IE +E+ QK +Y A L ++ +E+++H L

Sbjct: 805 NAVKPFVLRRLKEDVLQELPKIEHLQSSSELLPDQKRLYAAYLAKLREETLKHLDKDTLR 864

Query: 833 RRKIEILSGITRLRQICDTPRLFMD-YDGESGKLESLRQLLTQIKENGHRALIFSQFRGM 891  
 + KI IL+G+TRLRQIC+ P Lf+D Y G S KLE L +L + + G R LIFSQF M

Sbjct: 865 KNKIRILAGLTRLRQICNHPALFVDYKGSACKLEQLLDILEECRSTGKRILIFSQFTKM 924

50 Query: 892 LDIAEREMVAMGLTTYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADT 951  
 L I RE+ + + + G+TP+ ER E+ FN G D FLISLKAGG GLNLTGADT

Sbjct: 925 LSIIGRELNRQAIPYFYLDGNTPSQERVELCNRFNEGEGDLFLISLKAGGTGLNLTGADT 984

55 Query: 952 VVLIDLWNNPAVEMQAISRAHRLGQKENVEVYRLITRGITIEEKILEMQETKKHLVITVLD 1011  
 V+L DLWNNPAVE QA RA+R+GQK V+V +L+ GTIEEK+ E+QE+KKHL+ V++

Sbjct: 985 VILYDLWNNPAVEQQAADRAYRMGQKNTVQVIKLVAHGTTIEEKMHELQESKKHLIAEVIE 1044

Query: 1012 -GNETHASMSVDDIREIL 1028  
 G E +S++ ++IR+IL

60 Sbjct: 1045 PGEEKLSSITEEIRDIL 1062

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4823> which encodes the amino acid sequence <SEQ ID 4824>. Analysis of this protein sequence reveals the following:

Possible site: 26

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

65

-1738-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3909(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 674/1031 (65%), Positives = 834/1031 (80%), Gaps = 2/1031 (0%)

10 Query: 1 MSRMIPGRIRNQGIELYEQGLVSLISQEGNLLKAKVGDCQIEYSLVTEETKCSCDFFARK 60  
 M+R+IPGR+RN+GI+LYEQGLVS +L+ +V Q++Y E+ C CD F K  
 Sbjct: 2 MARLIPGRVRNEGILKLYEQGLVSPQDDNKGILQIEVETYQVQYGADDEDITCQCDTFHMK 61

15 Query: 61 GYCQHLLAALEHFLKNDPEKGAILSKVQVQESQOETKKKTSFSGSVFLDSLIIINEDDTIKY 120  
 YC+H+AA+E+FLKND +GK L ++ Q + ++ TTK TSFGS+FLDSL +NEDD++KY  
 Sbjct: 62 HYCKHIAAVEYFLKNDQKGLFLKQLTNQTKIKETTKKMTSFGSLFLDSLAMNEDDSVKY 121

20 Query: 121 QLSAQGEQNPYANDIWWTLKIRRLPDDRSYVIRDIKAFINTVRKEAYYQIGKQYFETLSL 180  
 +LSA G ++P+++D WW+LKI RLPDDRSYVIRDIK FL ++KE +YQIGK YFE LS  
 Sbjct: 122 RLSALGSRSPFSSDYWWSLKINRLPDDRSYVIRDIKGFLLQLIKKEGFYQIGKNYFEQLSW 181

25 Query: 181 IQFDETSQELIEFLWRLIPSHSSKIDLEFILPNQGRHLSLTRGFFEEGVTLMALENFSF 240  
 +QFD +SQ LIEFLWRL S + K D E I PN RHL L GFFEEG+ + +L +F+F  
 Sbjct: 182 LQFDPSSQALIEFLWRLA-SDTDKGDNENIFPNHARHLRLPSGFFEEGIIHYLTSLYDFTF 240

30 Query: 241 ESDFHQFNHLYFKELEGEDHLYQFKVIVHRQSIELEIKEKDLKPLFANSYLFYRDTFYHL 300  
 E ++HL+ + LE E LY+FKV VHR+SIEL+I EK+++ LF N YL Y+DTFYHL  
 Sbjct: 241 EGPSQTYHHLFVRSLEABAGLYEFKVEVHRKSIELQIAEKNVQYLFNDNDYLLYQDTFYHL 300

35 Query: 301 NLKQEKMTAIRSLPIEGDLAKHIFDLDDQDKLAAHLDFKEIGLVDAPRSFSIHDFKV 360  
 LKQ KMV AIRSLPIE DLAKHIFDLDD KLAA L DFK+IGLV+AP+SF+I DF+V  
 Sbjct: 301 TLKQKRMVQAIRSLPIEADLAKHIFDLDDHAKLAASLSDFKQIGLVEAPKSFAIRDFFV 360

40 Query: 361 NFEFDINSQNEILLQMVFDYGNLDTVHNROELEQLTFASHFKHEEKVFKLLEKYGFAPHF 420  
 F+FD+ +++EI Q++FDYGN V ++ LE L FASH K EEK+ + L +GF+P F  
 Sbjct: 361 TFQFDLLNRDEISQMLFDYGN-YQVSDKASLEALPFASHLKKEEKINRSLAFGFSPQF 419

45 Query: 421 STSHPAYSAQELYDFYTYMLPQFKKMGTVSLSAKLESYRLIERPQIDIEAKGSLLDISFD 480  
 + SA+ELY F+ +P F+++G V+LS +++ ++ E P+I I LLDISFD  
 Sbjct: 420 YSKRLTSAKELYTFFETVPCFERLGNVALSTAIQALQVKEMPKEAIRRNQGLLDISFD 479

50 Query: 481 FSDLLENDVDQALVALFDNPNPYFVNKSGQLVIFDEETKKVSATLQGLRARRAKNGHIELD 540  
 FS ++END+DQA+ ALF NNPYFV+++GQLV+FD+ET+KVS +LQ LRAR+ KNGH++LD  
 Sbjct: 480 FSTIIEENDIDQAVTALFQNNPYFVSQTGQLVVFDEETQKVSLSLQELRARQLKNGHLQLD 539

55 Query: 541 NIAAFQSELSEFANQDNVSFSQHIFYQLIEDLRHPEKFKIPGLSVSASLRDYQLTGVRWLSM 600  
 I A Q+S+LF +V FS+ +L L+HPE F I L V A +RDYQ GV+WLSM  
 Sbjct: 540 GIRALQVSKLFEGMTSVHFSKELEELAYHLQHPETFSIKPLPVKAQMRDYQRNGVQWLSM 599

60 Query: 601 LDHYGFAGILADDMGLGKTLQTSIFSLSTKLTRDSRVLILSPSSLIYNWQDEFHKFAPDVD 660  
 L+HYGF GILADDMGLGKTLQ++FL++ L DS+VLILSPSSLIYNW DE KF P +D  
 Sbjct: 600 LNHYGFAGILADDMGLGKTLQTLAFLASHLKSDSKVLILSPSSLIYNWFECQKFTPQLD 659

65 Query: 661 VAVAYGSKIRRDIIAERHQVIITSYSSFRQDFETYSEGNYDYILDEAQVMKNAQTKIA 720  
 V V+YG K RD+II E HQ+ ITSYSSFRQDFETY +DYILDEAQV+KNAQTKI+  
 Sbjct: 660 VVVSYGLKQIRDQIIEGHQITITSYSSFRQDFETYQAFHYDYILDEAQVIKNAQTKIS 719

Query: 721 HSLRSFEVKNCFALSGTPIENKLEIWSIFQIILPGLLPKGKKEFLKLNPKQVARYIKPFV 780  
 H LR+F NCFALSGTPIENK+LEIWSIFQI+LPGLLP KKEFLKL +QV+RYIKPFV  
 Sbjct: 720 HCLRAFNTANCFALSGTPIENKMLEIWSIFQIVLPGLLPKKEFLKLTAEQVSRYIKPFV 779

Query: 781 MRRRKEEVLPELPLDIEMNPNEMTDSQKVIYLAQLRQIQESIQHSSDADLNRRKIEILS 840  
 MRR+KE+VLELPLDIE+NY NEMTD QK IYLAQLRQ+Q+ I++SSD D++R+KIEILS  
 Sbjct: 780 MRRKEDVLELPLDIEINYSNEMTDEQKAIYLAQLRQMDQIRNSSDVIDISRQKIEILS 839

65 Query: 841 GITRLRQICDTPRLFMDYDGESGKLESRLQLLTQIKENGHRALIFSQFRGMLDIAEREMV 900  
 GITRLRQICDTP LFM DY G+SGKL+SLR LLTQIKENGHRALIFSQFRGMLD+A++EM

-1739-

Sbjct: 840 GITRIRQICDTPSLFMDYQKGSKLDSLRIILLTQIKENGHRALIFSQFRGMLDLAQEMT 899

Query: 901 AMGLITYKITGSTPANERHEMTRAFNAGSKDAFLISLKAGGVGLNLTGADTVVILDLWWN 960  
A+GLT+Y++TGSTPANER EMTRAFN GSKDAFLISLKAGGVG+NLTGADTV+LIDLWWN

5 Sbjct: 900 ALGLTSYQMTGSTPANERQEMTRAFNNGSKDAFLISLKAGGVGINLTGADTVILIDLWWN 959

Query: 961 PAVEMQAISRAHRLGQKENVEVYRLITRGTIEEKILEMQETKKHLVTTVLDGNETHASMS 1020  
PAVEMQAISRA+R+GQKENVEVYRLITRGTIEEKILE+QE+K++LVTTVLDGNE+ ASMS

10 Sbjct: 960 PAVEMQAISRAYRIGQKENVEVYRLITRGTIEEKILELQESKRNLVTTVLDGNESRASMS 1019

Query: 1021 VDDIREILGVS 1031  
+++I+EILG++

Sbjct: 1020 IEEIKEILGLN 1030

15 SEQ ID 4822 (GBS369) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 5; MW 120kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 6; MW 142kDa).

The GBS369-GST fusion product was purified (Figure 215, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 303), which confirmed that the protein is immunoaccessible  
20 on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1562

A DNA sequence (GBSx1654) was identified in *S.agalactiae* <SEQ ID 4825> which encodes the amino  
25 acid sequence <SEQ ID 4826>. Analysis of this protein sequence reveals the following:

Possible site: 41  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.3391(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

35 There is also homology to SEQ ID 1034:

Identities = 34/38 (89%), Positives = 37/38 (96%)

Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRTACQFS 38  
+EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRTAC+ S

40 Sbjct: 1 LEKEAKQMIDLKRNLFKIDVRAQKDEEKVFMRTACRQS 38

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

### Example 1563

45 A DNA sequence (GBSx1656) was identified in *S.agalactiae* <SEQ ID 4827> which encodes the amino acid sequence <SEQ ID 4828>. This protein is predicted to be phosphoglycerate dehydrogenase (era2). Analysis of this protein sequence reveals the following:

Possible site: 31  
>>> Seems to have no N-terminal signal sequence

50

-1740-

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3709(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA88823 GB:AB016077 phosphoglycerate dehydrogenase

[Streptococcus mutans]

Identities = 377/436 (86%), Positives = 414/436 (94%)

Query: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNKRFSLIDTGG 60  
 M LPTVAIVGRPNVGKS LFNRIAGERISIVEDVEGVTRDRIYT EWLNR+FS+IDTGG  
 Sbjct: 1 MALPTVAIVGRPNVGKSALFNRIAGERISIVEDVEGVTRDRIYTKAEWLNKRFSLIDTGG 60

Query: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVVS GKEGVTDADEYVSRIYKTNKPVILAVNK 120  
 IDVDVAPFMEQIKHQADIAMTEADVIVFVVS KEG+TDADEYV++ILY+T+KPVILAVNK  
 Sbjct: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVVSKEGITDADEYVAKILYRTHKPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIIRFSLI 180  
 VDNPEMR+ IYDFY+LGLGDPYP+SS HGIGTGD+LDAIV+NLP E + E+ DII+FSLI  
 Sbjct: 121 VDNPEMRSAIYDFYALGLGDPYPVSSAHGIGTGDVLDLDAIVDNLPAAEQEESDIIKFSLI 180

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNFVDSQGEYTMIDTAGMRKSGKVY 240  
 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT F D +GQE+TMIDTAGMRKSGKVY  
 Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTFTDEEGQEFTMIDTAGMRKSGKVY 240

Query: 241 ENTEKYSVMRSMRAIDRSDDVLMVINAEEGIREYDKRIAGFAHETGKGIIVVNKWDITIE 300  
 ENTEKYSVMR+MRAIDRS+VLMV+NAEEGIREYDKRIAGFAHE GKGI++VVNKWD I+  
 Sbjct: 241 ENTEKYSVMRAMRAIDRSDIVLMVINAEEGIREYDKRIAGFAHEAGKGIIVVNKWDIAIK 300

Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLPMIKRISESQNKRIIPSAVLN 360  
 KDN TV+QWE DIRDNFQ++ YAPI+FVSA TKQRLHKLPD+IK++S+SQN RIPS+VLN  
 Sbjct: 301 KDNRTVAQWETDIRDNFQYIPYAPIFVSAVTKQRLHKLPDVIKQVSQSQNTRIIPSSVLN 360

Query: 361 DVIMDAIAINPTPTDKGKRLKIFYATQVAVKPPTFVVFVNEEELMHFSYLRFLNQIREA 420  
 DV+MDA+AINPTPTDKGKRLKIFYATQV+VKPPTFV+VNEEELMHFSYLRFLNQIR+A  
 Sbjct: 361 DVVMDAIAINPTPTDKGKRLKIFYATQVSVKPPTFVIFVNEEELMHFSYLRFLNQIRQA 420

Query: 421 FVFEGTPIINLIARKRK 436  
 FVFEGTPI LIARKRK  
 Sbjct: 421 FVFEGTPIRLIARKRK 436

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4829> which encodes the amino acid sequence <SEQ ID 4830>. Analysis of this protein sequence reveals the following:

Possible site: 46  
 >>> Seems to have no N-terminal signal sequence

## ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3463(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 403/436 (92%), Positives = 422/436 (96%)

Query: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYTTGEWLNKRFSLIDTGG 60  
 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIY TGEWLNK+FSLIDTGG  
 Sbjct: 1 MVLPTVAIVGRPNVGKSTLFNRIAGERISIVEDVEGVTRDRIYATGEWLNKRFSLIDTGG 60

Query: 61 IDVDVAPFMEQIKHQADIAMTEADVIVFVVS GKEGVTDADEYVSRIYKTNKPVILAVNK 120  
 IDVDVAPFMEQIKHQ IAM EADVIVFVVS GKEGVTDADEYVS+ILY+TN PVILAVNK  
 Sbjct: 61 IDVDVAPFMEQIKHQADIAMEADVIVFVVS GKEGVTDADEYVSKILYRINTPVILAVNK 120

Query: 121 VDNPEMRNDIYDFYSLGLGDPYPLSSVHGIGTGDILDAIVENLPVEEENENPDIIIRFSLI 180

-1741-

VDNPEMRNDIYDFYSLGLGDPYP+SSVHGIGTGD+LDAIVENLPVEE EN DIIRFSLI  
 Sbjct: 121 VDNPEMRNDIYDFYSLGLGDPYPVSSVHGIGTGDVLDIAIVENLPVEEAEENDDIIRFSLI 180

Query: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTNFVDSQGEYTMIDTAGMRKSGKVY 240  
 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDT+F D+ GQE+TMIDTAGMRKSGK+Y  
 Sbjct: 181 GRPNVGKSSLINAILGEDRVIASPVAGTTRDAIDTHFTDADGQEF+TMIDTAGMRKSGKIY 240

Query: 241 ENTEKYSVMRSMRAIDRSDDVLMVINAEEGIREYDKRIAGFAHETGKGIIIVVNKWDITIE 300  
 ENTEKYSVMR+MRAIDRSDDVLMVINAEEGIREYDKRIAGFAHE GKG+IIIVVNKWDIT+  
 Sbjct: 241 ENTEKYSVMRAMRAIDRSDDVLMVINAEEGIREYDKRIAGFAHEAGKGMIIIVVNKWDITID 300

Query: 301 KDNHTVSQWEADIRDNFQFLSYAPIIFVSAETKQRLHKLPMIKRISSEONKRIPSAVLN 360  
 KDNHTV++WEADIRD FQFL+YAPIIFVSA TKQRL+KLDP+IKRISSEONKRIPSAVLN  
 Sbjct: 301 KDNHTVAKWEADIRDQFQFLTYAPIIFVSALTQRLNKLPLDIKRISSEONKRIPSAVLN 360

Query: 361 DVIMDAIAINPTPTDKGRLKIFYATQVAVKPPTFVVFVNEEELMHFSYLRFLNQIREA 420  
 DVIMDAIAINPTPTDKGRLKIFYATQV+VKPPTFVVFVNEEELMHFSYLRFLNQIR A  
 Sbjct: 361 DVIMDAIAINPTPTDKGRLKIFYATQVSVKPPTFVVFVNEEELMHFSYLRFLNQIRAA 420

Query: 421 FVFEGTPIINLIARKRK 436  
 F FEGTPI+LIARKRK  
 Sbjct: 421 FTFEGTPIHLIARKRK 436

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
 vaccines or diagnostics.

### Example 1564

A DNA sequence (GBSx1657) was identified in *S.agalactiae* <SEQ ID 4831> which encodes the amino acid sequence <SEQ ID 4832>. Analysis of this protein sequence reveals the following:

Possible site: 51  
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----  
 bacterial cytoplasm --- Certainty=0.2734(Affirmative) < succ>  
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00359 GB:AF008220 DnaI [Bacillus subtilis]  
 Identities = 105/313 (33%), Positives = 191/313 (60%), Gaps = 17/313 (5%)

Query: 1 MKSVGQALENQGRVP--RNTNDELIQMILADAQVAEFIKTHQ--LSQREINISMSKFNQF 56  
 M+ +G++L+ P + +++ + ++ D V F+K ++ + Q+ I S++K ++  
 Sbjct: 1 MEPIGRSLQGVTRPDFQKRLEQMKEKVMKDQDVQAFLKENEEVIDQKMIEKSLNKLIEY 60

Query: 57 LIERQK-----FKNKDSQYIAKGYEPILVMNEGYADVSYLE--TRELIEAQKKQAISDRI 109  
 IE+ K ++++ + +GY P LV+N D+ Y E + ++ QKKQ +  
 Sbjct: 61 -IEQSKNCSYCSSEDENCNNLLEGYHPKLVNNGRSIDIEYYECPVKRKLDDQKKQ--QSLM 117

Query: 110 NLVNLPKSYRNIRMTDFDINNESRMKAMSQLLDFVETYPYNSH-KGLYLYGDMGVGKSYL 168  
 + + + DI++ SR+ + DF+++Y KGLYLYG GVGK+++  
 Sbjct: 118 KSMYIQDQLLGATFQQVDISDPSRLAMFQHVITDFLKSINETGKGKGLYLYGKFGVGKTFM 177

Query: 169 MAAMARELSERKGVSTTLHFPSFAIDVKNAISSGTVKDEIDAVKSVFILILDIGAEQA 228  
 +AA+A EL+E++ S+ +++ P F ++KN++ T++++++ VK+ P+L+LDDIGAE  
 Sbjct: 178 LAAIANELAEKE-YSSMIVYVPEFVRELKNSLQDQTLLEKLNVMKTTTPVLMDDIGAESM 236

Query: 229 TSWVRDEILQVILQHRMLEELPTFFTSNYSFNDLERKWA-NIKGSDETQAKRVMERVRY 287  
 TSWVRDE++ +LQHRM ++LPTFF+SN+S ++L+ + + +G E +A R+MER+ Y  
 Sbjct: 237 TSWVRDEIVIGTLQHRMSQQLPTFFSSNFSPELKHHTYSQRGEKEEVKAARLMERILY 296

Query: 288 LAIEFHLEGNRR 300  
 LA L+G NRR

-1742-

Sbjct: 297 LAAPIRLDGENRR 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4833> which encodes the amino acid sequence <SEQ ID 4834>. Analysis of this protein sequence reveals the following:

5      Possible site: 19  
       >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----  
 10            bacterial cytoplasm --- Certainty=0.1944(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 228/300 (76%), Positives = 264/300 (88%)

15      Query: 1    MKSVGQALENQGRVPRNTNDELIQMILADAQVAEFIKTHQLSQREINISMSKFNQFLIER 60  
              M+ +G+ +    G+ R +D+LIQ ILAD +VA FI H LSQ +IN+S+SKFNQFL+ER  
       Sbjct: 1    MEKIGETMAKLGQNTRVNSDQLIQITILADPEVASFISQHHLSQEIQINLSLSKFNQFLVER 60

20      Query: 61   QKFKNKDSQYIAKGYEPILVMNEGYADVSYLETRELEIAQKKQAISDRINLVNLPKSYRN 120  
              QK++ KD   YIAKGY+PIL MNEGYADVSYLET+EL+EAQK+ AIS+RI LV+LPKSYR+  
       Sbjct: 61   QKYQLKDPSYIAKGYQPILAMNEGYADVSYLETKEQVEAQQAISERIQLVSLPKSYRH 120

25      Query: 121   IRMTDFDINNESRMKAMSQQLDFVETYPYNSHKGGLYLYGDMGVGKSYLMAAMARELSERK 180  
              I ++D D+NN SRM+A S +LDFVE YPS    KGLYLYGDMG+GKSYL+AAMA ELSE+K  
       Sbjct: 121   IHLSDIDVNNASRMEAFSAILDVFEQYPSAEQKGLYLYGDMGIGKSYLLAAMAHELSEKK 180

30      Query: 181   GVSTLLHFPSFAIDVKNAISSGTVKDEIDAVKSVPIILDDIGAEQATSWVRDEILQVI 240  
              GVSTLLHFPSFAIDVKNAIS+G+VK+EIDAVK+VP+LILDDIGAEQATSWVRDE+LQVI  
       Sbjct: 181   GVSTLLHFPSFAIDVKNAISNGSVKEEIDAVKNVPVILDDIGAEQATSWVRDEVLQVI 240

35      Query: 241   LQHRMLEELPTFFTSNYSFNDLERKWKANIKGSDETWQAKRVMERVRYLAIEFHLEGPNRR 300  
              LQ+RMLEELPTFFTSNYSF DLERKWA IKGSDETWQAKRVMERVRYLA EPHLEG NRR  
       Sbjct: 241   LQYRMLEELPTFFTSNYSFADLERKWATIKGSDETWQAKRVMERVRYLAREFHLEGANRR 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

#### Example 1565

40      A DNA sequence (GBSx1658) was identified in *S.galactiae* <SEQ ID 4835> which encodes the amino acid sequence <SEQ ID 4836>. Analysis of this protein sequence reveals the following:

      Possible site: 37  
       >>> Seems to have no N-terminal signal sequence

45      ----- Final Results -----  
              bacterial cytoplasm --- Certainty=0.2660(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
              bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

50      A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4837> which encodes the amino acid sequence <SEQ ID 4838>. Analysis of this protein sequence reveals the following:

55      Possible site: 35  
       >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----  
              bacterial cytoplasm --- Certainty=0.2135(Affirmative) < succ>  
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1743-

bacterial outside --- Certainty=0.0000 (Not Clear) &lt; succ&gt;

An alignment of the GAS and GBS proteins is shown below.

Identities = 217/391 (55%), Positives = 309/391 (78%)

5  
Query: 1 MMSPIDEFTYIKQNKIVYDSNSLIQLYFPIMGSDAMALYDYFVHFFDDGIRRHKFSEVLN 60  
MM PID FTY+K+K+ DS +LIQLYFPI+GSDA+++Y YF+HFFDDG++RHKFS++LN  
Sbjct: 1 MMKPIDTFTYLKRNKVTLDSTLIQLYFPIIGSDAVSIYQYFIHFFDDGLQRHKFSDILN 60

10  
Query: 61 HLQYGMMPRFQDALVMLTALDILLTVYQATGTLYLVKLNQAMSNEFLSNPIYRRLLEKRIGE 120  
HLQ+GM RF+DAL +LTA++L++VYQ + TYL+ L+Q +S +LF +P Y RLE++IGE  
Sbjct: 61 HLQFGMKRFEDALAILTAMELVSVYQLSDTYLITLHQPLSRDLFFQHPAYSRLLEQKIGE 120

15  
Query: 121 VAVAELDMKIPKNARDISKKFTDVFSDLGQPKQEVNRSKNVFDLESFKRLMMRDGLRFNN 180  
VAV+EL + +P AR+ISK+F+D+F G + + FDL SF++LM+RDGL+F +  
Sbjct: 121 VAVSELQVTVPSQARNISKRFSDIFGVQGDLTNVPQKPQKNFDLSSFQQLMVRDGLQFED 180

20  
Query: 181 EKDDVLGIYSVSELYHLNWDYTYQLAKQTAINGMIAPQRMKVQONEGQHIKDNQSFTNNE 240  
+ D++ +YS++E Y + W+DTYQ+AK TA+NG I P+R+ ++N+ ++F+ E  
Sbjct: 181 NQKDIISLYSIAEQYDMTWFDTYQIAKATAVNGKIRPERLLAKKNQSMTPSKENFSQAE 240

25  
Query: 241 KVILRESKNDLSALVFLEKIKRSKAVTTSGEKTLLEDLAKMNFLEVINVMVLYTLNKT 300  
++ILRE+K DSALVFLEKIK++R+A T E+ LL+ LAKMNFLE+VINVMVLYT NKT  
Sbjct: 241 QIILREAKQDSALVFLEKIKARRATITKDERILLQTLAKMNFLEDDVINVMVLYTFNKT 300

30  
Query: 301 SANLNKAYIMKVANDFAFQNVMTAEAVLKIRDFSDQKVRTKTETKKKQSNVPEWSNP 360  
SANL K+Y++K+ANDFA+Q V TAE+A++ +R F+D++ R +++ K QSNVP+WSNP 360  
Sbjct: 301 SANLQKSYVLKMANDFAYQKVSTAEAAIVVLRFTDRQSRQSKVKTQSNVPKWSNP 360

30  
Query: 361 KDEVSPEKEIELEQFKTDALKRLERLGKDG 391  
++ S E++ +L+QFK ALKRLE LGK G+  
Sbjct: 361 QETTSQEEQAKLDQFKQAALKRLENLKGKG 391

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1566

A DNA sequence (GBSx1659) was identified in *S. agalactiae* <SEQ ID 4839> which encodes the amino acid sequence <SEQ ID 4840>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4485 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06865 GB:AP001517 unknown conserved protein [Bacillus halodurans]

Identities = 80/150 (53%), Positives = 115/150 (76%)

50  
Query: 1 MRCPKCGYNKSSVVDSDRQAEETGTTIRRRRECEKCGNRFTTTFERLEELPLLVIKKGDTREQ 60  
MRCP C +N + V+DSR A EG +IRRRRECE C +RFTTTFE +EE+PL+V+KKGDT++  
Sbjct: 1 MRCPACHHNGTRVLDSRPAHEGRSIRRRRECESCNHRFTTTFEMIEEVPLIVVKKGDTREQ 60

55  
Query: 61 FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDVSSITIGNLVMDLAELD 120  
FS DKIL G+I++ +KRPV E+E + +ER++R + ++EV S IG LVM+ LA +D  
Sbjct: 61 FSSDKILRGLIRACEKRPVPLETLEGIVNEVERELRGQKNEVDSKEIGELVMERLANVD 120

60  
Query: 121 EITYVRFASVYKSFKDVEIEELLQITKR 150  
++ YVRFASVY+ FKD++ + L+++ +R  
Sbjct: 121 DVAYVRFASVYRQFKDINVFIQELKELMER 150

-1744-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4841> which encodes the amino acid sequence <SEQ ID 4842>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10                bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 131/155 (84%), Positives = 143/155 (91%)

15 Query: 1  MRCPCGYNKSSVVDSDRQAEAGTTIRRRRECEKCGNRFTTTFERLEELPLLVIKKDGTREQ 60
      +RCPKC Y+KSSVVDSDRQAE+G TIRRRRECE+C RFTTTFER+EELPLLVIKKDGTREQ
      Sbjct: 1  VRCPCKNYHKSSVVDSDRQAEAGNTIRRRRECEQCHTRFTTTFERVEELPLLVIKKDGTREQ 60

20 Query: 61  FSRDKILNGIIQSAQKRPVSSEDIENCILRIERKIRSEYEDEVSSITIGNLVMDLAELD 120
      FSRDKILNG++QSAQKRPVSS DIEN I RIE+++R+ YE+EVSS  IGNLVMDLAELD
      Sbjct: 61  FSRDKILNGVVQSAQKRPVSSTDIENVISRIEQEVRTTYENEVSSTAIGNLVMDLAELD 120

      Query: 121 EITYVRFASVYKSFKDVDEIEELLQQITKRVRSKK 155
      EITYVRFASVYKSFKDVDEIEELLQQIT RVR KK
25      Sbjct: 121 EITYVRFASVYKSFKDVDEIEELLQQITNRVRGKK 155

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1567

30 A DNA sequence (GBSx1660) was identified in *S.agalactiae* <SEQ ID 4843> which encodes the amino acid sequence <SEQ ID 4844>. This protein is predicted to be CsrS (mtrB). Analysis of this protein sequence reveals the following:

```

Possible site: 35
35 >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -11.30    Transmembrane    22 - 38 ( 18 - 43)
      INTEGRAL    Likelihood = -9.66     Transmembrane    189 - 205 ( 187 - 212)

----- Final Results -----
                bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
40                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2109> which encodes the amino acid sequence <SEQ ID 2110>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 35
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -6.32     Transmembrane    196 - 212 ( 189 - 214)

----- Final Results -----
50                bacterial membrane --- Certainty=0.3527(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

55 Identities = 248/501 (49%), Positives = 363/501 (71%), Gaps = 4/501 (0%)

```

-1745-

Query: 1 MKNKKDQFIGVKQPLSKKLSQLVFIFFSLFTVFSVLVYTSATRYVLHREKINVGRSLEK 60  
 M+N+K + K L K+LS + F+LFF +F+ F+++ Y+S ++L +EK +V +++  
 Sbjct: 1 MENQKQKQKKYKNSLPKRLSNIFVLFCCIFSAFTLIAYSSSTNYFLLKKEKQSVFQAVNI 60

5 Query: 61 TRVRLSQANSSLTSDILEILYNQVFADDIYPHKRQNGIVRTGESIDSILYVNQEMTLYD 120  
 RVRLS+ +S+ T +++ E+LY ++ + ++R+ I + L NQ++ +Y+  
 Sbjct: 61 VRVRLSEVDSNFTLENLAEVLYKNDKTHLRIDDRKGSRVIRSERDITNTLDANQDIYVYN 120

10 Query: 121 VNRKPVFST-LRTGMPITIGKSMGKVIISKVADM-EGFVGTKAIYSQKTGQLLGYVQIFYN 178  
 ++++ +F+T P + +G+V + D GF T+ +YS +TG+ +GYVQ+F++  
 Sbjct: 121 IDKQMIFTTDNEESSPGLHGPIGRVYHDHIEDQYRGFSMTQKVYSNRGKFGVGVQVFHD 180

15 Query: 179 LGRYYSMRQNIIVFLIMMEVLGTVLALVVINSATKRIVRPVKNLHDLMHQISENPSNLEI 238  
 LG YY +R ++ +L+++E+ GT LA ++I T+R ++P+ NLH++M ISENPNL +  
 Sbjct: 181 LGNYVIRARLLFWLLVVELFGTSLAYLIILITTRRFLKPLHNLHEVMRNISENPNNLNL 240

20 Query: 239 RSKVRSEDEIGELSRIFDGMLDQLEDYTRRQSQFISDVSHELRTPVAVVKGHIGLLQRWG 298  
 RS + S DEI ELS IFD MLD+LE +T+ QS+FISDVSHELRTPVA++KGHIGLLQRWG  
 Sbjct: 241 RSDISSGDEIEELSVIFDNMLDKLETHTKLQSRFISDVSHELRTPVAIKGHIGLLQRWG 300

25 Query: 299 KDDPEILEESLAAAYHEADRMSLMINDMLNMIRVQGSLELHQDEVTDLSSSISVVIENFR 358  
 KDD +ILEESL A HEADRM++MINDML+MIRVQGS E HQ+++T L SI V+ NFR  
 Sbjct: 301 KDDSDILEESLTATAHEADRMAIMINDMLDMIRVQGSFEGHQNDMTVLEDSIETVVGNGFR 360

30 Query: 359 ILREDFQFIFENNISDIVGKIYKIHFEQALMILIDNAIKYSPSYKEVSVVLSVDNDFAT 418  
 +LREDF F +++ + +IYK HFEQALMILIDNA+KYS K++++ LSV  
 Sbjct: 361 VLREDFIFTWQSENPKTI-ARIYKNHFEQALMILIDNAVKYRKEKKIATNLSVTGKQEA 419

35 Query: 419 VV-VKDKGEGISDEIDIEFIDRFYRTDKSRNRESTQAGLGIGLSVFKQIMDAYHLKVDIK 477  
 +V V+DKGEGIS EDIE IF+RFYRTDKSRNR STQAGLGIGLS+ KQI+D YHL++ ++  
 Sbjct: 420 IVRVQDKGEGISKEDIEHIFERFYRTDKSRNRTSTQAGLGIGLSILKQIVDGYHLQMKVE 479

Query: 478 SELNQGTEFIVRIPIKKFEET 498  
 SELN+G+ FI+ IP+ + +E+  
 Sbjct: 480 SELNEGSVFILHIPLAQSKES 500

A related GBS gene <SEQ ID 8845> and protein <SEQ ID 8846> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 5  
 SRCFLG: 0  
 McG: Length of UR: 5  
 Peak Value of UR: 0.74  
 Net Charge of CR: 2

45 McG: Discrim Score: -10.19  
 GvH: Signal Score (-7.5): -3.66  
 Possible site: 35  
 >>> Seems to have no N-terminal signal sequence  
 Amino Acid Composition: calculated from 1  
 ALOM program count: 2 value: -11.30 threshold: 0.0

50 INTEGRAL Likelihood = -11.30 Transmembrane 22 - 38 ( 18 - 43)  
 INTEGRAL Likelihood = -9.66 Transmembrane 189 - 205 ( 187 - 212)  
 PERIPHERAL Likelihood = 2.86 405  
 modified ALOM score: 2.76  
 icml HYPID: 7 CFP: 0.552

55 \*\*\* Reasoning Step: 3

----- Final Results -----  
 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60

-1746-

SEQ ID 8846 (GBS321) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 6; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 80 (lane 2; MW 58.7kDa).

GBS321-GST was purified as shown in Figure 220, lane 3.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1568

A DNA sequence (GBSx1661) was identified in *S.agalactiae* <SEQ ID 4845> which encodes the amino acid sequence <SEQ ID 4846>. This protein is predicted to be CsrR (trcR). Analysis of this protein  
10 sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2649(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3259> which encodes the amino acid  
20 sequence <SEQ ID 3260>. Analysis of this protein sequence reveals the following:

Possible site: 61  
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.3226(Affirmative) < succ>  
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 193/229 (84%), Positives = 211/229 (91%), Gaps = 1/229 (0%)

Query: 1 MGKKILIIIEDEKNLARFVSLELLHEGYDVVETNGREGLDTALEKDFDLILLDLMLPEMD 60  
M KKILIIIEDEKNLARFVSLEL HEGY+V+VE NGREGL+TALEK+FDLILLDLMLPEMD  
Sbjct: 1 MTKKILIIIEDEKNLARFVSLELQHEGYEVIVEVNGREGLETALEKEFDLILLDLMLPEMD 60

35 Query: 61 GFEITRRLQAEKTTYIMMTARDSVMDIVAGLDRGADDYIVKPFATIEELLARVRAIFRRQ 120  
GFE+TRRLQ EKTYYIMMTARDS+MD+VAGLDRGADDYIVKPFATIEELLAR+RAIFRRQ  
Sbjct: 61 GFEVTRRLQTEKTTYIMMTARDSIMDVVAGLDRGADDYIVKPFATIEELLARIRAIFFRRQ 120

40 Query: 121 EIETKTKKEKGDGSGSFRDLSLNTHNRSAMRGDEEISLTKREFDLLNVLMTNMNRVMTREEL 180  
+IE++ K+ G +RDL LN NRS RGD+EISLTKRE+DLIN+LMTNMNRVMTREEL  
Sbjct: 121 DIESE-KKVPSQGIYRDLVLNPQNRSVNRGDDEISLTKREYDLLNIILMTNMNRVMTREEL 179

45 Query: 181 LEHVWKYDVAETNVVDVYIRYLRGKIDIPGRESYIQTVRGMGYVIREK 229  
L +VWKYD A ETNVVDVYIRYLRGKIDIPG+ESYIQTVRGMGYVIREK  
Sbjct: 180 LSNVWKYDEAVETNVVDVYIRYLRGKIDIPGKESYIQTVRGMGYVIREK 228

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### 50 Example 1569

A DNA sequence (GBSx1662) was identified in *S.agalactiae* <SEQ ID 4847> which encodes the amino acid sequence <SEQ ID 4848>. Analysis of this protein sequence reveals the following:

-1747-

Possible site: 60

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

----- Final Results -----

5                   bacterial cytoplasm --- Certainty=0.3864(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10           >GP:AAG32547 GB:U12643 YlbN-like hypothetical protein [Streptococcus gordonii]  
             Identities = 91/174 (52%), Positives = 133/174 (76%), Gaps = 3/174 (1%)

Query: 3    LTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVYEDGLYLDDYNMAYDIT 62  
           + EI+K+P+GL F+KK+D+ E L ER++EI+D+ DI SG YEDGLY LDY ++Y IT  
 15   Sbjct: 4   IQEIRKNPDGLAFEKKLDLAEELKERNAEILDVQDIVASGRAQYEDGLYFLDYELSYTIT 63

Query: 63   LPSSSRMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNILL 122  
           L SSRSM+PV E +NE+F+E V++ +E++DQ+LVL +E +IN+ ESVDNILL  
 20   Sbjct: 64   LASSRSMFPVERKESYLVNEIFMEDGQVAS-QEMIDQDLVLPINGEINVAESVADNILL 122

Query: 123   NIPLRLVLADEVGVADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175  
           NIPL+VL AA+E G + +G++W +MTE Y++ QA++KE+++PFA L+G+FD  
 25   Sbjct: 123   NIPLKVLTAEEAGSDLP-TGRDWQVMTEDDYQKYQAEKKEENSPFAGLQGLFD 175

25   A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4849> which encodes the amino acid  
      sequence <SEQ ID 4850>. Analysis of this protein sequence reveals the following:

Possible site: 42

&gt;&gt;&gt; Seems to have no N-terminal signal sequence

30           ----- Final Results -----

                  bacterial cytoplasm --- Certainty=0.3032(Affirmative) < succ>  
                   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>  
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35   An alignment of the GAS and GBS proteins is shown below.

Identities = 86/175 (49%), Positives = 135/175 (77%)

Query: 1    MLLTEIKKSPEGLYFDKKIDIKESLMERHSEIMDISDIQVSGHVYEDGLYLDDYNMAYD 60  
           + ++EI+K P+GL FD+ D+K L+ER +I+DI ++ G+V Y+ GLYLDDY ++Y+  
 40   Sbjct: 3   LAISEIRKHPDGLSFDRLCDVKSMLLERDQQTIDIKAVKAVGNVRYDKGLYLDDYQLSYE 62

Query: 61   ITLPSSSRMKPVVLSEKQTINEVFIEAENVSTKKELVDQELVLILEEDDINLEESVIDNI 120  
           + LPSSSRM PV LSE Q I E+FIEA +++ KKELV+ LVL+L++D INLEES++DNI  
 45   Sbjct: 63   VILPSSSRMVPVCLSEVQHIQELFIEATDLADKKELVEDNLVLVLDKDAINLEESIVDNI 122

Query: 121   LLNIPLRLVLADEVGVADLSGKNWSLMTEKQYEEKQAKEKEKSNPFAALEGMFD 175  
           LL IP++VL +E + +G+NW+++TE+ Y+ + +++++NPFA+L+G+FD  
 50   Sbjct: 123   LLAIPVQVLTEEEKKSKELPAGQNNAVLTEDDYQCLKEEKQKKNPFAALQGLFD 177

50   Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for  
      vaccines or diagnostics.

**Example 1570**

55   A DNA sequence (GBSx1663) was identified in *S.galactiae* <SEQ ID 4851> which encodes the amino  
      acid sequence <SEQ ID 4852>. This protein is predicted to be heat shock protein (htpX). Analysis of this  
      protein sequence reveals the following:

Possible site: 25

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

INTEGRAL   Likelihood = -11.30   Transmembrane   195 - 211 ( 190 - 221)

-1748-

INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 ( 31 - 62)  
 INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 ( 153 - 174)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5522(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAB70525 GB:AF017421 putative heat shock protein HtpX  
 [Streptococcus gordonii]  
 Identities = 220/297 (74%), Positives = 261/297 (87%), Gaps = 1/297 (0%)

15 Query: 1 MLYQQIASNKRKTIVLLIVFFCLLAAIGAAGVYLVLGSYQFGLVLALIIGVIYAVSMIFQ 60  
 ML++QIA+NKR+T LL+ FF LLA IGAA GYL + S G+++A IIG+IYA++MIFQ  
 Sbjct: 1 MLFEQIAANKRRTWFLLVAFALLALIGAAAGYLMWNSPLGGVIAFIIGLIYAITMIFQ 60

20 Query: 61 STNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPEN 120  
 ST VVMSMN AR+V+E EAP +HIV+DMAM+AQIPMPRV+IVEDDS NAFATGS PEN 120  
 Sbjct: 61 STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSNPEN 120

25 Query: 121 AVAATTGLLAVMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180  
 AVAATTGLL +MNRREELEGVIGHEVSHIRNYDIRISTIAVALASA+T+ISS+ RM++YG  
 Sbjct: 121 AVAATTGLLRLMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

30 Query: 181 GRRRRDDDDREDGG-NILVLIFSILSLILAPLAASLVQLAISRQREYLADASSVELTRNPQ 239  
 GRRRR+D +D G +L+L+FS++++ILAPLAA+LVQLAISRQRE+LADASSVELTRNPQ  
 Sbjct: 181 GRRRRNDRDDDSGLGLMLVFSLIAIILAPLAATLVQLAISRQREFLADASSVELTRNPQ 240

30 Query: 240 GMISALEKLD RSEPMGHPVDDASAALYINDPTKKEGLKSLFYTHPPADIADRIERLRHM 296  
 GMI AL+KLD SEPM VDDASAALYI+DP KK GL+ LFYTHPPI++R+ERLR M  
 Sbjct: 241 GMIRALQKLDNSEPMHRHVDDASAALYISDPKKKGGLQLFYTHPPISERVERLRKM 297

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 4853> which encodes the amino acid  
 sequence <SEQ ID 4854>. Analysis of this protein sequence reveals the following:

Possible site: 31

&gt;&gt;&gt; Seems to have a cleavable N-term signal seq.

40 INTEGRAL Likelihood = -9.77 Transmembrane 197 - 213 ( 192 - 223)  
 INTEGRAL Likelihood = -8.33 Transmembrane 43 - 59 ( 33 - 61)  
 INTEGRAL Likelihood = -3.82 Transmembrane 153 - 169 ( 153 - 174)

----- Final Results -----

45 bacterial membrane --- Certainty=0.4906(Affirmative) < succ>  
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>  
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

50 >GP:AAB70525 GB:AF017421 putative heat shock protein HtpX [Streptococcus gordonii]  
 Identities = 208/298 (69%), Positives = 257/298 (85%), Gaps = 1/298 (0%)

55 Query: 61 STSLVMSMNNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIIEDPSLNAFATGSSPQNA 120  
 ST +VMSMN AR+V+E+EAP +HIV+DMAMVAQIPMPRV+I+ED S NAFATGS+P+NA  
 Sbjct: 61 STEVVMSMNGARQVSEQEAPELYHIVQDMAMVAQIPMPRVYIVEDDSPNAFATGSNPEN 120

60 Query: 121 AVAATTGLLVEMNRREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGGRMLWYG 180  
 AVAATTGLL +MNRREELEGVIGHE+SHIRNYDIRISTIAVALASA+T+ISS+ GRM+WYG  
 Sbjct: 121 AVAATTGLLRLMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAITMISSVAGRMMWYG 180

Query: 181 GGSRRQRDDGDDVLRITITLLSLLSLLAPLVASLIQLAISRQREYLADASSVELTRNP 240  
 GG RR+ D DD L ++ L+ SL++++LAPL A+L+QLAISRQRE+LADASSVELTRNP

-1749-

Sbjct: 181 GG-RRRNDRDDSGGLGLMLVFSLIAITLAPLAATLVQLAISRQREFLADASSVELTRNP 239

Query: 241 QGMKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSSLFSTHPPIEERIERLKNM 298

QGM+AL+KL S+PM VDDASAALYI++P+KK LF THPPI ER+ERL+ M

Sbjct: 240 QGMIRALQKLDNSEPMHRHVDASAALYISDPKKKGGLQKLFYTHPPISEVERLERLKM 297

An alignment of the GAS and GBS proteins is shown below.

Identities = 233/298 (78%), Positives = 262/298 (87%), Gaps = 2/298 (0%)

10 Query: 1 MLYQQIASNKRKTVVLLIVFFCLLAAIGAAGVYLVLGSYQFGLVLALIIGVIYAVSMIFQ 60  
 MLYQQI+ NK++TVVLL+ FF LLA IGA+ GYL+L +Y GLVLAL+IGVIYA SMIFQ  
 Sbjct: 1 MLYQQISQNKQRTVVLLVGFALLALIGASAGYLLLDNYAMGLVLALVIGVIYATSMIFQ 60

15 Query: 61 STNVVMSMNNAREVTEDEAPNYFHIVEDMAMIAQIPMPRVFIVEDDSLNAFATGSKPEN 120  
 ST++VMSMNNAREVTE EAP +FHIVEDMAM+AQIPMPRVFI+ED SLNAFATGS P+NA  
 Sbjct: 61 STSLVMSMNNAREVTEKEAPGFFHIVEDMAMVAQIPMPRVFIIEDPSLNAFATGSSPQNA 120

20 Query: 121 AVAATTGLLAVMNRREELEGVIGHEVSHIRNYDIRISTIAVALASAVTLISSIGSRMLFYG 180  
 AVAATTGLL VMNRREELEGVIGHE+SHIRNYDIRISTIAVALASAVT+ISSIG RML+YG  
 Sbjct: 121 AVAATTGLLEVMNRREELEGVIGHEISHIRNYDIRISTIAVALASAVTVISSIGGRMLWYG 180

25 Query: 181 GG--RRDDDDREDGGNILLVLFISLILAPLAASLVQLAISRQREYLADASSVELTRNP 238  
 GG R+RDD +D I+ L+ S+LSL+LAPL ASL+QLAISRQREYLADASSVELTRNP  
 Sbjct: 181 GGSRRQRDDGDDVLRITLLLSLLSLLLAPLVLASLIQLAISRQREYLADASSVELTRNP 240

Query: 239 QGMISALEKLDRESEPMGHPVDDASAALYINDPTKKEGLKSLFYTHPPIADRIERLRHM 296  
 QGMI ALEKL S+PM HPVDDASAALYIN+P KK SLF THPPI +RIERL++M  
 Sbjct: 241 QGMKALEKLQLSQPMKHPVDDASAALYINEPRKKRSFSSSLFSTHPPIEERIERLKNM 298

A related GBS gene <SEQ ID 8847> and protein <SEQ ID 8848> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 9.61

GvH: Signal Score (-7.5): -0.97

Possible site: 25

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 3 value: -11.30 threshold: 0.0

INTEGRAL Likelihood = -11.30 Transmembrane 195 - 211 ( 190 - 221)

INTEGRAL Likelihood = -11.09 Transmembrane 43 - 59 ( 31 - 62)

INTEGRAL Likelihood = -3.61 Transmembrane 153 - 169 ( 153 - 174)

PERIPHERAL Likelihood = 5.89 87

modified ALOM score: 2.76

\*\*\* Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.5522(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

73.8/88.3% over 296aa

imported

SP|O30795| PUTATIVE HEAT SHOCK PROTEIN HTPX. Insert characterized

GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus gordonii} Insert characterized

PIR|T48855|T48855 probable heat shock protein HtpX - Streptococcus gordonii Insert characterized

ORF02338(301 - 1188 of 1488)

SP|O30795|HTPX\_STRGC(1 - 297 of 297) PUTATIVE HEAT SHOCK PROTEIN

HTPX.GP|2407215|gb|AAB70525.1||AF017421 putative heat shock protein HtpX {Streptococcus gordonii}|PIR|T48855|T48855 probable heat shock protein HtpX [imported] - Streptococcus gordonii

```
%Match = 44.0
%Identity = 73.7 %Similarity = 88.2
Matches = 219 Mismatches = 34 Conservative Sub.s = 43
```

[illegible]

35 SEQ ID 8848 (GBS179) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 175 (lane 11; MW 58kDa).

GBS179-GST was purified as shown in Figure 227, lane 5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

### Example 1571

40 A DNA sequence (GBSx1665) was identified in *S.agalactiae* <SEQ ID 4855> which encodes the amino acid sequence <SEQ ID 4856>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL      Likelihood =-15.44    Transmembrane      4 - 20 ( 1 - 27)

----- Final Results -----
      bacterial membrane --- Certainty=0.7177(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG23700 GB:AF017421 LemA-like protein [Streptococcus gordonii]  
Identities = 124/182 (68%). Positives = 152/182 (83%)

55 Query: 1 MGMTILIAIIALFVITWLIVAYNSLVRSRMHTKESWSQIDVQLKRNDLIPNLIIETVKGYA 60  
M +I IA+I + V+++I YNSLVR+RM T+E+WSQIDVQLKRNDL+PNLIETVKGY  
Sbjct: 1 MSFIITIAIVIVIVLFLVISVYNSLVRARMQTQEAWSQIDVQLKRNDLIPNLIIETVKGYG 60  
  
Query: 61 AYEGKTLEKIAELRAQVAKANTPAEAMTASNELTROLSSILAVAENYPDLKANNSFVKILO 120